

AUS DEM LEHRSTUHL  
FÜR EXPERIMENTELLE MEDIZIN UND THERAPIEVERFAHREN  
PROF. DR. CHRISTOPH KLEIN  
DER FAKULTÄT FÜR MEDIZIN  
DER UNIVERSITÄT REGENSBURG

**Kombinierte Genom- und Transkriptomanalyse von einzelnen  
disseminierten Tumorzellen aus dem Knochenmark von  
Prostatakarzinompatienten**

Inaugural-Dissertation  
zur Erlangung des Doktorgrades  
der Medizin

der Fakultät für Medizin  
der Universität Regensburg

vorgelegt von  
Bernhard Braun

2015



AUS DEM LEHRSTUHL  
FÜR EXPERIMENTELLE MEDIZIN UND THERAPIEVERFAHREN  
PROF. DR. CHRISTOPH KLEIN  
DER FAKULTÄT FÜR MEDIZIN  
DER UNIVERSITÄT REGENSBURG

**Kombinierte Genom- und Transkriptomanalyse von einzelnen  
disseminierten Tumorzellen aus dem Knochenmark von  
Prostatakarzinompatienten**

Inaugural-Dissertation  
zur Erlangung des Doktorgrades  
der Medizin

der Fakultät für Medizin  
der Universität Regensburg

vorgelegt von  
Bernhard Braun

2015

Die vorliegende Arbeit entstand im Zeitraum vom Oktober 2009 bis September 2013 unter der Anleitung von Herrn Prof. Dr. Christoph Klein am Lehrstuhl für Experimentelle Medizin und Therapieverfahren am Klinikum der Universität Regensburg.

Dekan: Prof. Dr. Dr. Torsten E. Reichert

1. Berichterstatter: Prof. Dr. Christoph A. Klein

2. Berichterstatter: Prof. Dr. Maximilian Burger

Tag der mündlichen Prüfung: 03.12.2015





# Inhaltsverzeichnis

<b>Abkürzungsverzeichnis</b>	<b>6</b>
<b>1. Hintergrund und Fragestellung</b>	<b>8</b>
<b>2. Material und Methoden</b>	<b>10</b>
2.1. Patienten	10
2.2. Aufbereitung der Knochenmarksproben	11
2.3. Immunfluoreszenzfärbung	12
2.4. Untersuchung der Knochenmarksproben und Isolierung von Einzelzellen und Zellpools	13
2.5. Isolierung und Amplifikation des Einzelzelltranskriptoms	14
2.6. Isolierung und Amplifikation des Einzelzellgenoms und komparative genomische Hybridisierung	15
2.7. Optimierung der Bedingungen für die genspezifische PCR	16
2.8. Genspezifische PCR zur Detektierung ausgewählter Transkripte in Patientenproben	20
2.9. Microarray basierte Transkriptomanalyse	20
2.10. Statistische Auswertung	21
<b>3. Ergebnisse</b>	<b>22</b>
3.1. Nachweis von EpCAM <sup>+</sup> Zellen in den Knochenmarksproben und Korrelation mit klinischen Parametern	22
3.2. EpCAM Färbung in Knochenmarksproben von Prostatakarzinom-Patienten und der Kontrollgruppe	23
3.3. Isolation, Genom- und Transkriptomamplifikation von EpCAM <sup>+</sup> Einzelzellen	28
3.4. Gezielte Transkriptomanalyse zur Identifikation einer DCC Signatur bei Patienten im M0 Stadium	29
3.5. PCR Profiling von EpCAM <sup>+</sup> Zellen aus dem Knochenmark von metastasierten Patienten	36
3.6. Genexpressionsprofile und phänotypische Heterogenität	36
3.7. Analyse der genomischen Aberrationen	38
3.8. Transkriptomanalyse von EpCAM <sup>+</sup> Einzelzellen mittels Microarrays	41
<b>4. Diskussion</b>	<b>44</b>

<b>5. Zusammenfassung</b>	<b>50</b>
<b>6. Literaturverzeichnis</b>	<b>51</b>
<b>7. Wissenschaftliche Originalarbeit und Supplementary Material</b>	<b>57</b>
<b>8. Danksagung</b>	<b>104</b>
<b>9. Erklärung</b>	<b>105</b>

## Abkürzungsverzeichnis

APC	Allophycocyanin
aCGH	Array basierte komparative genomische Hybridisierung
bp	Basenpaar
BSA	bovines Serumalbumin
CD	cluster of differentiation
cDNA	komplementäre Desoxyribonukleinsäure
CGH	komparative genomische Hybridisierung
CK	Zytokeratin
CTC	circulating tumor cell
Cy3	Cyanin 3
DAPI	4',6'-Diamidino-2'-Phenylindol-Dihydrochlorid
dATP	2'-Desoxyadenosin-5'-triphosphat
DCC	disseminated cancer cell
dCTP	2'-Desoxycytidin-5'-triphosphat
gGTP	2'-Desoxyguanosin-5'-triphosphat
Dig	Digoxigenin
DNA	Desoxyribonukleinsäure
dNTP	2'-Desoxynukleosid-5'-triphosphat
DTT	Dithiothreitol
dTTP	2'-Desoxythymidin-5'-triphosphat
dUTP	2'-Desoxyuridin-5'-triphosphat
EDTA	Ethylendiamintetraessigsäure
EpCAM	epitheliales Zelladhäsionsmolekül
FACS	Durchflusszytometrie
g	Beschleunigung
gDNA	genomische Desoxyribonukleinsäure
GlyA	Glycophorin A
h	Stunde
KCl	Kaliumchlorid
KH <sub>2</sub> PO <sub>4</sub>	Kaliumdihydrogenphosphat
KM	Knochenmark
MACS	magnetisch aktivierte Zellseparation
MgCl <sub>2</sub>	Magnesiumchlorid
min	Minute
MNZ	mononukleäre Zelle

mRNA	Boten-RNA
PBL	periphere Blutlymphozyten
PBS	phosphatgepufferte Kochsalzlösung
PCa	Prostatakarzinom
PCR	Polymerase-Kettenreaktion
PE	Phycoerythrin
PNA	Poly-T-Peptid-Nukleinsäure
RE	Restriktionsendonuklease
RNA	Ribonukleinsäure
rpm	Umdrehungen pro Minute
SDS	Natriumdecosulfat
sec	Sekunde
SSC	Natriumcitrat-Kochsalzlösung
TdT	terminale Desoxyribonukleotidyltransferase
tRNA	Transfer-RNA
WGA	whole genome amplification
WTA	whole transcriptome amplification

## 1. Hintergrund und Fragestellung

Trotz erheblicher Fortschritte im Bereich der Tumorforschung ist die Entstehung von Metastasen immer noch wenig verstanden, obwohl sie für mehr als 90% der tumorassoziierten Todesfälle verantwortlich ist (Brabletz et al., 2013). Die ersten Stufen der Metastasenentstehung näher zu charakterisieren – d.h. bevor manifeste Läsionen mittels der verfügbaren Bildgebung festgestellt werden können – könnte einen guten Ansatz darstellen die Metastasierung in Zukunft zu verhindern. Nachdem sie kurz im Blut zirkulieren (hier werden sie als zirkulierende Tumorzellen (circulating tumor cells, CTCs) bezeichnet), siedeln sich die Tumorzellen in anderen Organen ab und werden dort disseminierte Tumorzellen (disseminated cancer cells, DCCs) genannt. Tumorzellen, die nach einer Operation mit kurativer Intention im Körper verbleiben, stellen offensichtlich die Gründerpopulation der später auswachsenden Metastasen dar und eignen sich somit als Ziel für systemische, adjuvante Therapien. Zahlreiche Studien belegen die prognostische Relevanz disseminierter Tumorzellen. DCCs werden üblicherweise in Lymphknoten und Knochenmark mittels histogenetischer Marker wie dem epithelialen Zelladhäsionsmolekül (EpCAM) oder Zytokeratinen (CK) entdeckt. Anders als das Vorhandensein von CTCs, die bisher meist nur in Patienten mit manifesten Metastasen detektiert und charakterisiert wurden (Allard et al., 2004), ist das Vorhandensein von DCCs in frühen, nicht-metastasierten Krankheitsstadien von eindeutiger Relevanz (Braun et al., 2005; Stoecklein et al., 2008; Weckermann et al., 2009).

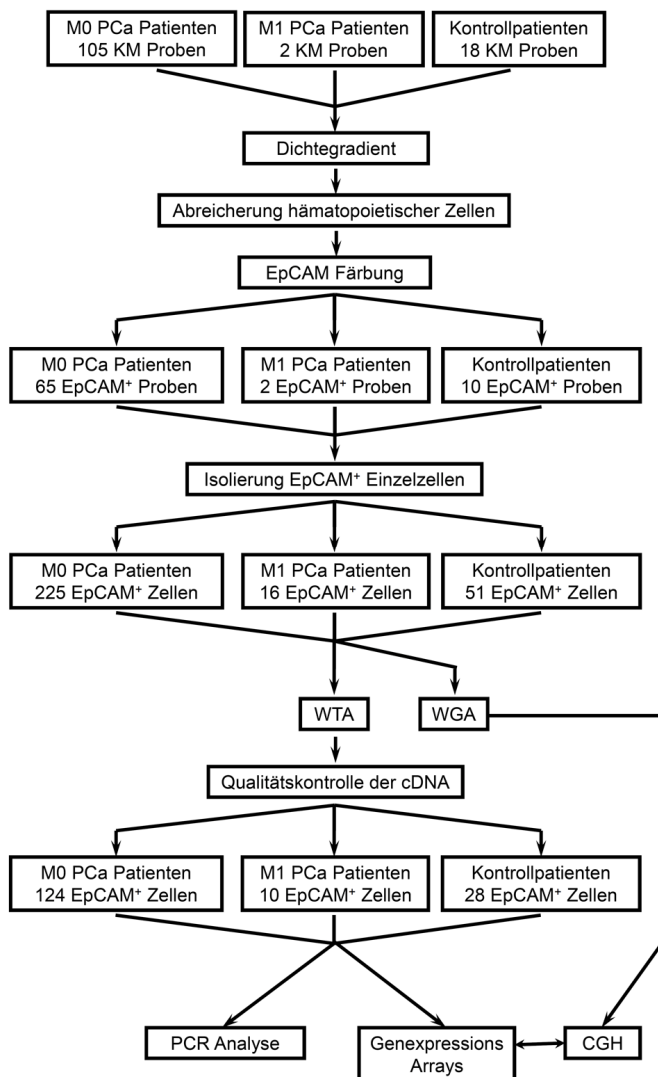
Das Vorhandensein CK-positiver ( $CK^+$ ) Zellen im Knochenmark von nicht-metastasierten Prostatakarzinom-Patienten zum Zeitpunkt der Prostatektomie ist mit einem kürzeren metastasenfreien Überleben assoziiert (Weckermann et al., 2009). Genetische Analysen der isolierten  $CK^+$  Zellen erbrachten eine ausgesprochene genetische Heterogenität der DCCs sowohl innerhalb als auch zwischen den einzelnen Patienten im M0-Stadium (Klein et al., 2002; Weckermann et al., 2009). Über phänotypische Eigenschaften der DCCs ist jedoch noch wenig bekannt. Überlebensdaten suggerieren das Vorhandensein von DCCs mit unterschiedlichen funktionellen Eigenschaften, da nicht alle Patienten mit DCCs letztendlich Metastasen entwickeln. Die Tatsache, dass auch einige Patienten ohne Nachweis CK-positiver Zellen im Knochenmark Metastasen entwickelten, spricht für das Vorhandensein von DCC-Subpopulationen, die mit den derzeitigen Markern nicht entdeckt werden. Dies trifft vor allem bei Prostatakarzinompatienten zu, da der PSA-Anstieg nach Prostatektomie (biochemisches Rezidiv) nicht mit steigender Anzahl CK-positiver Zellen im Knochenmark assoziiert ist (Weckermann et al., 2009). Um gezielte Therapien gegen diejenigen Zellen zu entwickeln, die letztlich das Metastasenwachstum bedingen, ist eine eingehende Charakterisierung der DCCs notwendig. Dies wird durch das seltene Vorkommen von DCCs im Knochenmark erschwert, die in nicht-metastasierten Tumorstadien nur in einer durchschnittlichen Häufigkeit von 1-2 DCCs pro  $10^6$  Knochenmarkszellen gefunden werden.

Eine Methode zur umfassenden Transkriptomanalyse von Einzelzellen mittels Microarray wurde in unserem Labor bereits etabliert (Hartmann et al., 2006). Da Zytokeratine intrazelluläre Proteine sind, erfordert eine umfassende Transkriptomanalyse von Einzelzellen die Verwendung eines Oberflächenmarkers, zum Beispiel EpCAM, für die Detektion vitaler DCCs. Das Vorhandensein EpCAM-positiver (EpCAM<sup>+</sup>) Zellen im Knochenmark wurde bereits im Zusammenhang mit der Progression des Prostatakarzinoms beschrieben (Morgan et al., 2009). Anders als CK<sup>+</sup> Zellen, die kaum im Knochenmark von gesunden Personen gefunden werden (Klein, 2003), werden schwach EpCAM<sup>+</sup> Zellen regelmäßig im Knochenmark gesunder Probanden detektiert (Klein et al., 2002; Choessel et al., 2004) und stören somit die molekulare Analyse von DCCs. Bevor also eine verlässliche Genom- und Transkriptomanalyse disseminierter Tumorzellen erfolgen kann, benötigt man eine Methode diese eindeutig zu identifizieren.

Das Ziel dieser Arbeit war es, EpCAM<sup>+</sup> Einzelzellen im Knochenmark von nicht-metastasierten Prostatakarzinompatienten und gesunden Probanden zu detektieren, zu isolieren und anschließend an ihnen eine gezielte Transkriptomanalyse durchzuführen. Hierzu wurde die Expression einer Reihe ausgewählter relevanter Marker in EpCAM<sup>+</sup> Einzelzellen aus dem Knochenmark von Prostatakarzinompatienten und gesunden Probanden analysiert. Im Einzelnen wurde die Expression von epithelialen (*EPCAM*, *KRT8*, *KRT18*, *KRT19*, *KRT5*, *KRT6*, *KRT14*), prostataassoziierten (*KLK3/PSA*, *AR*, *ERG*, *CD44*, *PROM1/CD133*), tumorspezifischen (*MAGEA2*, *MAGEA4*) und hämatopoetischen (*PTPRC/CD45*, *CD34*, *CD33*, *CD19*, *GPYC*, *SCL4A1*, *HBA2*, *TFRC/CD71*, *KIT/CD117*, *EPOR*, *NFE2*, *KLF1*) Transkripten untersucht. Das Ziel war hierbei war die Trennung von vermeintlich hämatopoetischen und epithelialen Zellen. Anschließend wurden ausgewählte Einzelzellen mittels komparativer genomischer Hybridisierung (CGH) und Genexpression Microarray Analyse einer umfassenden kombinierten Genom- und Transkriptomanalyse unterzogen.

## 2. Material und Methoden

Im Folgenden ein graphischer Überblick über die Vorgehensweise in der vorliegenden Arbeit:



**Abbildung 1.** Arbeitsablauf der vorliegenden Arbeit. Die Knochenmarksproben wurden nach Abreicherung hämatopoietischer Zellen auf das Vorhandensein von EpCAM<sup>+</sup> Einzelzellen untersucht. Aus jeder isolierten Einzelzelle, bzw. jedem Zellcluster/Zellpool wurde die gesamte mRNA und genomische DNA isoliert, aufgetrennt und jeweils amplifiziert, um anschließend den entsprechenden Untersuchungen zugeführt zu werden.

### 2.1 Patienten

Als Grundlage für die Verwendung der Patientenproben dient das Ethikvotum 07-079 der Universität Regensburg. Im Zeitraum zwischen Februar 2009 und November 2012 wurden 105 Knochenmarksaspirate von Prostatakarzinompatienten ohne manifeste Metastasen (M0) und 2 Aspirate von Patienten mit Metastasen (M1) gesammelt. Die Proben wurden jeweils vor Prostatektomie entnommen. Als Kontrollgruppe dienten 18 Knochenmarksaspirate von männlichen



Patienten ohne bekannte maligne Erkrankung, die im Rahmen einer unfallchirurgischen bzw. orthopädischen Operation entnommen wurden. Nach schriftlichem Einverständnis wurden die Proben durch beidseitige Aspiration am Beckenkamm gewonnen (Fehm et al., 2006) und innerhalb von 3 Stunden für die weitere Aufarbeitung an den Lehrstuhl für Experimentelle Medizin und Therapieverfahren gebracht. Die klinischen Charakteristika der Patienten sind in Tabelle 3 aufgeführt.

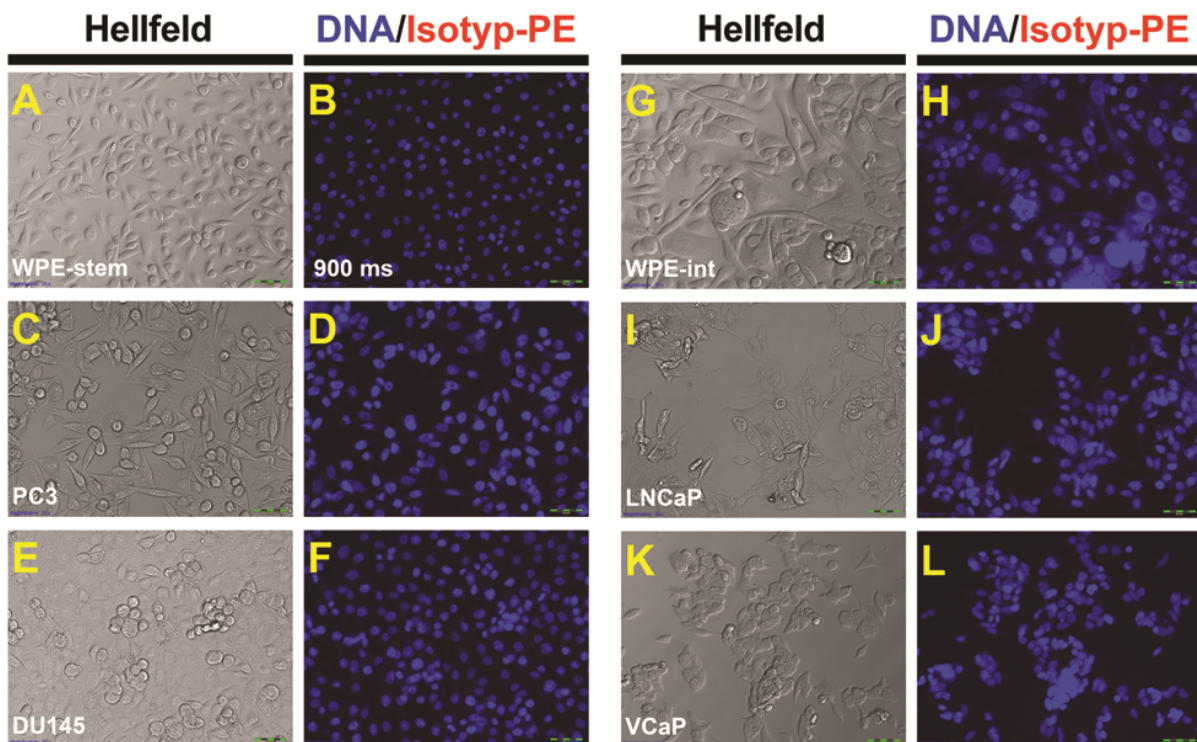
## **2.2 Aufbereitung der Knochenmarksproben**

Nach Ankunft der Proben im Labor wurde das KM-Volumen notiert und das Knochenmark in einem 50 ml Zentrifugenröhrchen mit Hank's Salzlösung (HSS; Biochrom AG) bis zu einem Endvolumen von 50 ml aufgefüllt. Die Probe wurde zunächst 10 min bei 170 x g bei 4°C zentrifugiert. Der Überstand (enthält Thrombozyten, Reste von Fetten und Serum) wurde entfernt und die ganze Prozedur noch einmal wiederholt. Danach wurde das Zellpellet in 9 ml Hank's Salzlösung resuspendiert und vorsichtig über 6 ml 65%-ige Percoll Lösung ( $\rho = 1,083 \text{ g/cm}^3$ ; GE Healthcare) geschichtet. Um Erythrozyten und Granulozyten zu entfernen wurde die Probe 20 min bei 1000 x g bei 4°C zentrifugiert. Anschließend wurde die Interphase mit den mononukleären Zellen (MNZ) vorsichtig entnommen und mit phosphatgepufferter Kochsalzlösung gewaschen, um verbliebene Percollreste zu entfernen. Die Suspension wurde 10 min bei 500 x g bei 4°C zentrifugiert und der Überstand verworfen. Die Anzahl der mononukleären Zellen und der verbliebenen Erythrozyten wurde mit Hilfe der Neubauer-Zählkammer bestimmt. Zur Anreicherung der Zellfraktion, die potentielle disseminierte Tumorzellen enthält, wurde mittels negativer immunomagnetischer Selektion ein Großteil der hämatopoetischen Zellen entfernt. Hierzu wurde das Zellpellet in MACS Puffer ( $90 \mu\text{l}/10^7 \text{ MNZ}$ ; 0,5% BSA, 2 mM EDTA in PBS) resuspendiert. Die Zellsuspension wurde anschließend mit APC-konjugierten Antikörpern gegen CD11b, welches von Monozyten, Granulozyten, Makrophagen und natürlichen Killerzellen exprimiert wird ( $10 \mu\text{l}/10^7 \text{ MNZ}$ ; die Konzentration der Antikörper war den Herstellerangaben nicht zu entnehmen, Miltenyi Biotec), CD33, welches von myeloischen Zellen exprimiert wird ( $5 \mu\text{l}/10^7 \text{ MNZ}$ ; Invitrogen), und CD45, ein Antigen, welches von einem Großteil der Leukozyten exprimiert wird ( $5 \mu\text{l}/10^7 \text{ MNZ}$ ; Invitrogen), inkubiert. Nach einer Inkubationszeit von 15 min bei 4°C wurde die Zellsuspension mit MACS Puffer gewaschen und in MACS Puffer resuspendiert ( $60 \mu\text{l}/10^7 \text{ Erythrozyten}$ ). Anschließend wurden anti-APC-Beads ( $20 \mu\text{l}/10^7 \text{ MNC}$ ; Miltenyi Biotec) und anti-CD235a-Beads (Glycophorin A, auf reifen erythrozytären Zellen exprimiert,  $20 \mu\text{l}/10^7 \text{ Erythrozyten}$ ; Miltenyi Biotec) hinzugefügt. Nach 15-minütiger Inkubation bei 4°C wurde die Suspension erneut mit MACS Puffer gewaschen. Das Pellet wurde in 2 ml MACS Puffer resuspendiert und über ein Zellsieb (Porengröße 40  $\mu\text{m}$ ) in ein neues Zentrifugenröhrchen filtriert. Die Suspension wurde auf die LS MACS Säule (Miltenyi Biotec) aufgetragen, die zuvor mit MACS Puffer gespült wurde. Das Eluat (Negativfraktion) wurde

gesammelt und auf Eis gekühlt und die Säule mit 9 ml MACS Puffer gespült. Die Zellzahl in der gesammelten Negativfraktion (11 ml) wurde dann mittels Neugebauer-Zählkammer bestimmt.

## **2.3 Immunfluoreszenzfärbung**

Im Durchschnitt wurden zwei Millionen Knochenmarkszellen in 100 µl Blocklösung [10% humanes AB Serum (Biotest) in PBS] resuspendiert und 15 Minuten auf einem Rollgerät bei 4°C inkubiert. Dann wurden 2 µl anti-EpCAM-PE Antikörper (HEA125; Miltenyi Biotec, die Antikörperkonzentration wurde vom Hersteller nicht angegeben) hinzugegeben und für weitere 15 min auf dem Rollgerät inkubiert. Nach der Inkubation wurde die Probe 5 min bei 500 x g zentrifugiert und das Zellpellet einmal mit PBS gewaschen. In 12 Fällen wurde eine Färbung mittels Isotyp Antikörper (Dako) unter gleichen Bedingungen durchgeführt (insgesamt 10 % aller untersuchten Knochenmarksproben). In den übrigen Fällen wurde keine Isotyp Kontrolle angefertigt, da nicht ausreichend Zellmaterial vorhanden war (Teile der gleichen Knochenmarksproben wurden für andere *in vivo* und *in vitro* Experimente benötigt). Alle Experimente, an denen Zelllinien beteiligt waren, enthielten auch eine Isotyp Kontrolle (Abb. 2). Die Zelllinien (WPE-Stem, DU145, PC3, WPE-Int, LNCaP und VCaP) wurden gemäß der Anweisungen von ATCC kultiviert. Nachdem eine Konfluenz von 80% erreicht war, wurde das Medium entfernt, die Zellschicht mittels PBS gewaschen und die Zellen anschließend in 4%-iger Formaldehydlösung fixiert. Danach wurden sie erneut mit PBS gewaschen. Die weiteren Schritte der Immunfluoreszenzfärbung wurden wie zuvor beschrieben durchgeführt. Abschließend wurden die gefärbten Zellen mittels einer DAPI Lösung (0,5 ng/µl in PBS) zur Kerngegenfärbung für 5 min inkubiert und dann mit frischem PBS gewaschen.



**Abbildung 2.** Photos der verwendeten epithelialen Prostatazelllinien, die mit dem Isotyp Antikörper inkubiert wurden. Der Zellname befindet sich in der unteren linken Ecke der Hellfeld-Abbildungen. Alle Photos wurden mit einer Belichtungszeit von 900 ms aufgenommen. Maßstab: 50  $\mu\text{m}$ . Es wurden keine unspezifischen Bindungen beobachtet.

## 2.4 Untersuchung der Knochenmarksproben und Isolierung von Einzelzellen und Zellpools

Jede Knochenmarkspröbe wurde manuell nach dem Vorhandensein von EpCAM<sup>+</sup> Zellen unter dem Fluoreszenzmikroskop (Olympus oder Zeiss) untersucht. Zur Zellisolierung war dieses mit einem Mikromanipulator (Patchman NP2, Eppendorf) und einer Pumpvorrichtung (CellTram, Eppendorf) ausgestattet. 30  $\mu\text{l}$  Zellsuspension mit ungefähr  $0,3 \times 10^6$  gefärbter KM-Zellen wurden in eine Kammer des 8-Kammer Objektträgers (Nunc) pipettiert, die bereits 170  $\mu\text{l}$  PBS enthielt. Einzelzellen, die eine intakte Morphologie aufwiesen und gleichzeitig als EpCAM<sup>+</sup> eingestuft wurden, wurden mittels der am Mikromanipulator angebrachten Glaskapillare isoliert. Nach dem Entfernen der Einzelzelle wurde diese in eine Kammer transferiert, die 200  $\mu\text{l}$  PBS enthielt, um per visueller Kontrolle sicherzustellen, dass auch nur eine einzelne Zelle isoliert wurde. Diese wurde dann mit Hilfe einer Mikropipette manuell in einem Mikroliter PBS abpipettiert und in ein kleines 0,2 ml Reaktionsgefäß überführt. Dieses enthielt 10 ng tRNA (Roche), die in 4  $\mu\text{l}$  Lysispuffer (mTRAP kit, Active Motif) gelöst war. Die Proben wurden dann unverzüglich bei  $-80^\circ\text{C}$  gelagert. Zusätzlich zu den Einzelzellen wurden auch Zellpools von etwa 2000 bis 3000 Zellen isoliert. Hierzu wurde 1  $\mu\text{l}$  der Zellsuspension abpipettiert und in ein kleines Reaktionsgefäß mit tRNA und Lysispuffer überführt. Als Negativkontrolle diente 1  $\mu\text{l}$  zellfreies PBS aus der Kammer, aus der die Einzelzellen

isoliert wurden. Die Negativkontrolle wurde beim Prozess der Transkriptomamplifikation (*whole transcriptome amplification* – WTA) mitgeführt.

In zehn Fällen wurden vor der Zellisolierung Photos der gefärbten Zellen mit Hilfe der Software cellSens (Olympus) oder AxioVison (Zeiss) gemacht. Da der Fluoreszenzfarbstoff Phycoerythrin sehr schnell ausbleicht, wurde darauf verzichtet von allen gefärbten Proben Photos anzufertigen. Die durchflusszytometrische Analyse der Knochenmarkszellen wurde mittels LSR II Apparat durchgeführt. Dieser war mit der Software FACS DIVA 5.03 (BD Bioscience) ausgestattet. Die Ergebnisse wurden von der Software FloJo 8.8.6 (Treestar) ausgewertet.

## 2.5 Isolierung und Amplifikation des Einzelzelltranskriptoms

Die Isolierung der mRNA aus den Einzelzellen, Transkription in cDNA und Amplifikation der cDNA wurde nach veröffentlichten Protokollen durchgeführt (Klein et al., 2002; Hartmann et al., 2006). Die isolierten Einzelzellen wurden in je ein Reaktionsgefäß transferiert, welches 6,4 µl Lysispuffer (Active Motif) mit 10 ng tRNA (Roche), 1 µg Protease (Active Motif) und 1 µl einer 37,5 µM Lösung mit biotinylierten oligo-dT Peptid-Nukleinsäuren (PNAs; Active Motif) enthielt. Für den proteolytischen Verdau wurden die Proben 10 min bei 45°C inkubiert, danach wurde die Protease durch Inkubation für 1 min bei 75°C deaktiviert. Für das Binden der PNAs an die poly-A Enden der mRNAs wurden die Proben anschließend 15 min bei 22°C gelagert. Die gebildeten PNA-mRNA Komplexe wurden mittels Streptavidin-konjugierter magnetischer Beads in einem Magnetfeld ausgefällt. Während des Ausfällens in einem Magnetständer wurden die Bead-Pellets mit 10 µl cDNA Waschpuffer 1 (50 mM Tris-HCl pH 8,3, 75 mM KCl, 10 mM DTT und 0,25% Igepal), 20 µl cDNA Waschpuffer 2 (50 mM Tris-HCl pH 8,3, 75 mM KCl, 10 mM DTT, 0,25% Tween-20) und erneut 20 µl cDNA Waschpuffer 1 gereinigt. Der Überstand mit darin enthaltener DNA wurde für die weitere DNA-Fällung und *whole genome amplification* (WGA) in ein Reaktionsgefäß überführt, welches 0,8 µl polymerisierter 0,25%-iger Polyacrylamid-Lösung enthielt.

Die cDNA Transkription wurde unter dauerhafter Rotation (verhindert das Absinken der Beads) für 45 min bei 44°C durchgeführt. Der Reaktion wurde ein Gemisch aus 0,5 mM äquimolaren dNTPs (GE Healthcare), 200 U Superscript II Reverse Transkriptase (Invitrogen), 30 µM CFL15CN8 Primer (C<sub>15</sub>GTCTAGAN<sub>8</sub>), 15 µM CFL15CT24 Primer (C<sub>15</sub>GTCTAGAT<sub>24</sub>VN; beide Metabion), 0,25% Igepal, 5 mM DTT (Invitrogen) und der geeignete RT Puffer vom Hersteller (Invitrogen), zu einem Endvolumen von 20 µl zugefügt. Vor Hinzugabe des Enzyms wurde das Reaktionsgemisch 10 min bei Raumtemperatur gelagert, damit sich die Primer anlagern konnten. Nach der cDNA Synthese wurden die Beads mit 20 µl Waschpuffer (50 mM KH<sub>2</sub>PO<sub>4</sub> pH 7, 1 mM DTT, 0,25% Igepal) gewaschen und in 10 µl Tailingpuffer (10 mM KH<sub>2</sub>PO<sub>4</sub> pH 7, 4 mM MgCl<sub>2</sub>, 0,1 mM DTT, 200 µM dGTP) resuspendiert. Das Reaktionsgemisch wurde mit 40 µl Mineralöl überlagert und die

cDNA Einzelstränge durch 5-minütige Inkubation bei 95°C von den Beads gelöst. Die Proben wurden anschließend auf Eis gekühlt. Das Hinzufügen von dGTPs an das 5' Ende der cDNA Einzelstränge wurde durch Hinzugeben von 10 U der Terminal-Deoxynukleotid-Transferase (TdT, USB-Affymetrix) und Inkubation der Reaktion für 60 min bei 37°C durchgeführt. Nach Inaktivierung der TdT (70°C, 5 min) wurden 35 µl WTA Gemisch 1 (4 µl Puffer 1 (Expand Long Template, Roche), 3% deionisiertes Formamid) zu jeder Probe hinzugefügt. Mittels Hinzugeben von 5,5 µl WTA Gemisch 2, welches 3,2 mM äquimolare dNTPs, 12 mM CP2 Primer (TCAGAATTCATGC<sub>15</sub>) und 7,5 U PolMix (Expand Long Template) enthielt, wurde eine *hot-start* PCR gestartet. Insgesamt wurden 40 PCR Zyklen in einem Thermoblock (MJ research PCR thermal cycler) durchgeführt: 20 Zyklen mit Denaturierungszeit 15 sec bei 94°C, Hybridisierungszeit 30 sec bei 65°C, Elongationszeit 2 min bei 68°C und 20 Zyklen mit einer Verlängerung der Elongationszeit von 10 sec sowie einer finalen Elongation von 7 min bei 68°C. Die Qualität der Amplifikation wurde mittels PCR auf die Haushaltsgene *ACTB*, *EEF1A1* und *GAPDH* überprüft (siehe auch folgende Abschnitte und Anhang Tabelle S1)

## **2.6 Isolierung und Amplifikation des Einzelzellgenoms und komparative genomische Hybridisierung**

Sechzig Mikroliter des zuvor gewonnen Überstandes mit darin enthaltener DNA und Polyacrylamid-Lösung wurden mit 120 µl 100%-igen eiskalten Ethanol gemischt und zur DNA-Fällung bei -20°C gelagert. Danach wurden die Reaktionsgefäße 45 min bei 20800 x *g* und 4°C zentrifugiert. Nach Entfernen des Überstandes wurde das Pellet mit 180 µl von 70%-igen eiskalten Ethanol im Thermomixer (Eppendorf) gewaschen (10 min, 18°C, 350 rpm). Danach wurden die Reaktionsgefäße bei Raumtemperatur für 10 min bei 20800 x *g* abzentrifugiert. Diese Reinigungsschritte wurden danach zweimal wiederholt. Nach Lufttrocknen des Pellets wurde es in 3,5 µl Wasser resuspendiert und erneut im Thermomixer (Eppendorf) inkubiert (18 h, 18°C, 350 rpm). Die anschließende Amplifikation des Einzelzellgenoms (WGA) wurde mittels Ampli1™ WGA Kit (Silicon Biosystems) durchgeführt. Die hier zugrunde liegende Methode wurde kürzlich veröffentlicht (Klein et al., 1999; Klein et al., 2002). Die komparative genomische Hybridisierung (CHG) mittels Metaphasen (cCGH, n=24 Proben) und Agilent Mikroarrays (aCGH, n=18 Proben) wurde gemäß veröffentlichter Protokolle durchgeführt (Klein et al., 1999; Klein et al., 2002; Czyz et al., 2014). Die komparativen genomischen Hybridisierungen wurden von Herrn Dr. Miodrag Guzvic durchgeführt. Die Rohdaten der Arrays sind auf der GEO Datenbank unter der Zugangsnummer GSE59631 zu finden.

## 2.7 Optimierung der Bedingungen für die genspezifische PCR

Die Primer für jedes ausgewählte Gen wurden anhand der cDNA Sequenz (Benson et al., 2010; Maglott et al., 2011) mittels dem Primer 3 Softwaretool ausgewählt (Rozen et al., 2000). Die Spezifität der Primer wurde noch einmal mit dem BLAST Softwaretool überprüft (Altschul et al., 1990). Hiernach konnten die verwendeten Primerpaare alle Isoformen der korrespondierenden Transkripte erkennen. Die einzige Ausnahme bildete *MAGEA2*, dessen Primer lediglich sechs der vorhandenen sieben Isoformen erkannten. Die Amplifikation der *MAGEA* Gene hat sich als sehr schwierig herausgestellt, weshalb wir bereits bekannte Primer verwendeten (Kufer et al., 2002; Mecklenburg et al., 2007)

Die WTAs von verschiedenen Einzelzellen aus ausgewählten Zelllinien (VCaP, PC3, A431 und BT20) wurden zur Optimierung der PCR Bedingungen für die Amplifikation der einzelnen Gene verwendet (Tabelle 1). Diese stammen aus einer Sammlung an Zelllinien WTAs, die in den vergangenen Jahren am Lehrstuhl für Experimentelle Medizin und Therapieverfahren etabliert wurde. Die Hybridisierungstemperaturen aller Primer wurden mittels Gradienten-PCR ermittelt, wobei jeweils Temperaturen im Bereich 5°C über und unter der errechneten Schmelztemperatur des Primerpaars getestet wurden. Die PCR Reaktion wurde in einem 10 µl enthaltenden Gemisch aus 0,1 µl WTA cDNA, 10 mM Tris-HCl, 50 mM KCl, 1 mM MgCl<sub>2</sub>, 0,1 mM äquimolares Gemisch aus dNTPs (GE Healthcare), 0,4 pM jedes Primers (Tabelle 1), 0,5 µg/µl an BSA (Roche) und 0,5 U Taq DNA Polymerase (Pan Biotech) durchgeführt. Das PCR Profil bestand aus folgenden Teilen: Erste Phase: Denaturierung bei 94°C über 2 min, Hybridisierung (Temperatur genspezifisch gewählt, siehe Tabelle 1) über 30 sec, und Elongation bei 72 °C für 2 min, insgesamt ein Zyklus; zweite Phase: Denaturierung bei 94°C für 15 sec, Hybridisierung über 30 sec, und Elongation bei 72 °C über 20 sec, insgesamt 14 Zyklen; dritte Phase: Denaturierung bei 94°C für 15 sec, Hybridisierung über 30 sec, und Elongation bei 72 °C über 30 sec, insgesamt 25 Zyklen; vierte Phase: finale Elongation bei 72°C über 2 min. Alle PCR Produkte wurden durch ein 3-%iges Agarosegel aufgetrennt und mittels Ethidiumbromid unter UV-Licht sichtbar gemacht. Um die korrekte Amplifikation eines Fragmentes zu beweisen wurden alle PCR Produkte mittels ausgewählter Restriktionsendonukleasen verdaut (NEB; siehe Abb. 3 und Tabelle 1).

Der Restriktionsverdau wurde in 30 µl eines wasserbasierten Reaktionsgemischs durchgeführt, welches 15 µl des primären PCR Produkts, 3 µl der entsprechenden Restriktionsendonuklease (Tabelle 1) und 3 µl des passenden Reaktionspuffers (NEB; NEBuffer 1 für Tsp509I, NEBuffer 3 für BsrI und NEBuffer 4 für Hpy188I, Hpy188III, HpyCH4IV, HpyCH4V, HaeIII, NlaIII und EarI) enthält. Für die Enzyme Hpy188III und NlaIII wurde das Reaktionsgemisch mit 10 mg/µl BSA (Roche) versetzt. Die Inkubationszeit für Hpy188I, Hpy188III, HpyCH4IV, HpyCH4V, EarI und NlaIII betrug 3 h bei 37°C, gefolgt von einer Inaktivierung für 20 min bei 65°C; für das Enzym

HaeIII 3 h Inkubation bei 37°C und Inaktivierung für 20 min bei 80°C, für Tsp509I 3 h Inkubation bei 65°C ohne folgende Inaktivierung und für BsrI 3h Inkubation bei 65°C und Inaktivierung für 20 min bei 80°C.

Zudem wurden alle Primer an humaner genomischer DNA (Invitrogen) getestet. Mit keinem der verwendeten Primer konnten die erwarteten Fragmente amplifiziert werden, was deren Spezifität für cDNA beweist.

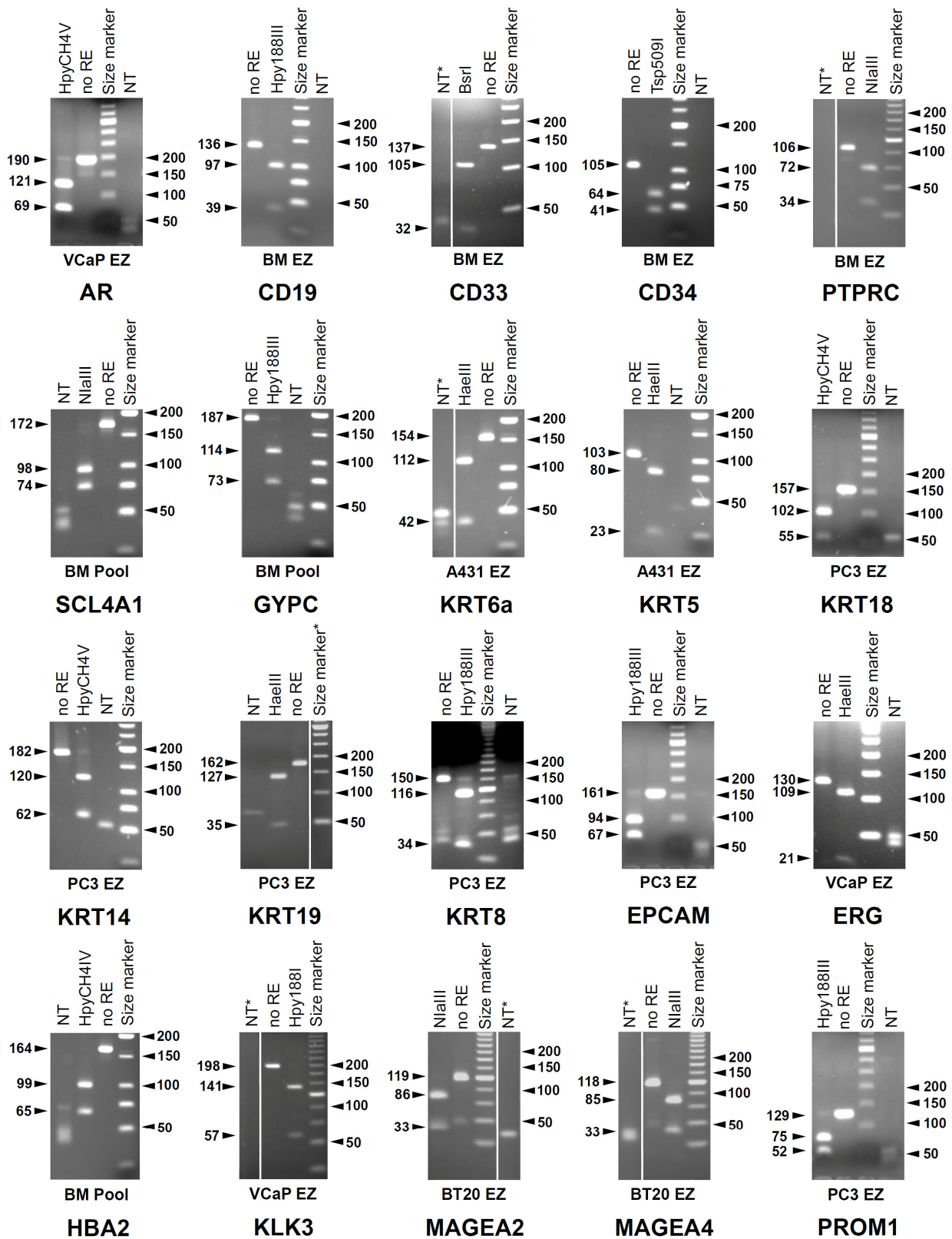
**Tabelle 1.** Liste der untersuchten Transkripte (GeneBank Name und gebräuchliches Synonym), Primersequenz, Hybridisierungstemperatur (Ta), Restriktionsendonukleasen und Fragmentgröße nach Restriktionsverdau der PCR Produkte. bp – Basenpaar; n.v. – nicht verfügbar.

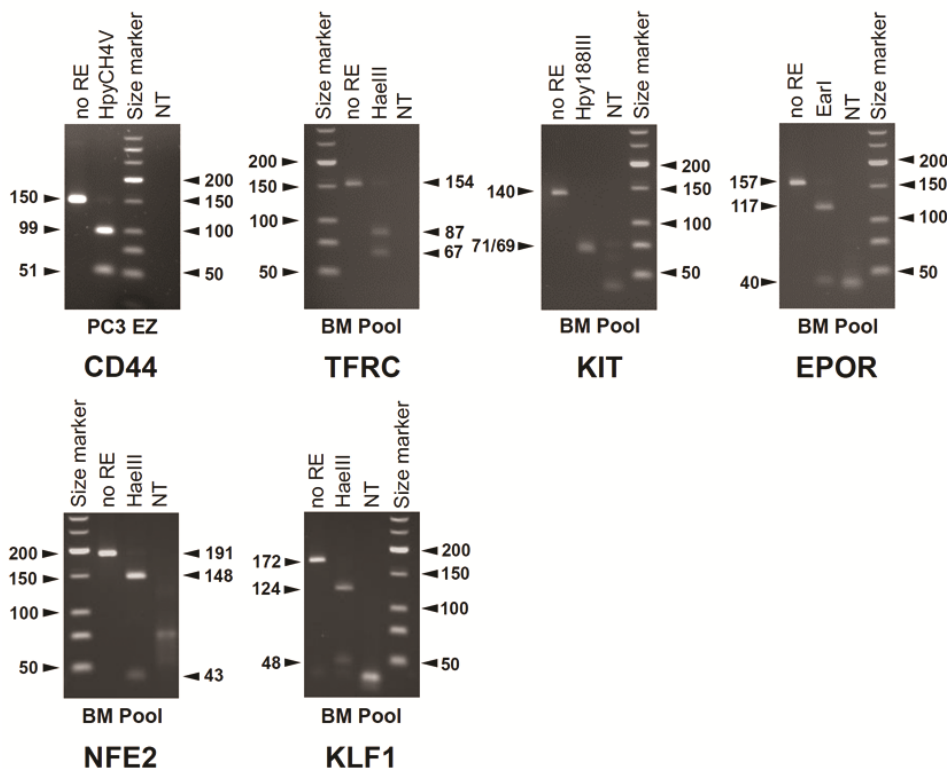
Gen	GenBank Nummer	Primer	Sequenz (5'-3')	Fragment (bp)	Ta (°C)	Restriktion senzym	Fragmente/ Verdau (bp)
<i>EPCAM</i>	NM_002354.2	Forward Reverse	AAAGTTTGC G GACTGCACTT AGCCACATCAGCTATGTCCA	161	55	Hpy188III	94 + 67
<i>KRT8</i> (CK8)	NM_002273.3	Forward Reverse	GCGTACAGAGATGGAGAACGA AGCTCCCGGATCTCCTCTT	150	58	Hpy188III	116 + 34
<i>KRT18</i> (CK18)	NM_000224.2	Forward Reverse	GGAAGATGGCGAGGACTTTA CCTGCTTCTGCTGGCTTAAT	157	55	HpyCH4V	102 + 55
<i>KRT19</i> (CK19)	NM_002276.4	Forward Reverse	AGCCGGACTGAAGAATTGAA TTCTGCCAGTGTGTCTTCCA	162	62	HaeIII	127 + 35
<i>KRT5</i> (CK5)	NM_000424.3	Forward Reverse	GGAGCTCATGAACACCAAGC TGGTCCAACCTCCTTCTCCAC	103	64	HaeIII	80 + 23
<i>KRT6a</i> (CK6a)	NM_005554.3	Forward Reverse	GTGAGGAGTGCAGGCTGAAT CCAAGACCACTGCCATAGGA	154	64	HaeIII	112 + 42
<i>KRT14</i> (CK14)	NM_000526.4	Forward Reverse	CGCAGTCATCCAGAGATGTG GGGATCTTCCAGTGGGATCT	182	62	HpyCH4V	120 + 62
<i>KLK3</i> (PSA)	NM_001648.2	Forward Reverse	TTGTCTTCCTCACCTGTCC TTTTGTTCTGATGCAGTGG	198	58	Hpy188I	141 + 57
<i>AR</i>	NM_000044.2	Forward Reverse	ATGTCCTGGAAGCCATTGAG TGAATGACAGCCATCTGGTC	190	58	HpyCH4V	121 + 69
<i>ERG</i>	NM_001136154.1	Forward Reverse	CCTTGCAAATCCAGGCAGT GTCGGGATCCGTCATCTTG	130	62	HaeIII	109 + 21
<i>CD44</i>	NM_000610.3	Forward Reverse	CGGACACCATGGACAAGTTT CCGTCCGAGAGATGCTGTAG	150	58	HpyCH4V	99 + 51
<i>PROM1</i> (CD133)	NM_006017.2	Forward Reverse	ATGACAAGCCCATCACAACA TTGCTCCTGGATTTGGAAG	129	55	Hpy188III	77 + 52
<i>MAGEA2</i>	NM_175743.1	Forward Reverse	CATTGAAGGAGAAGATCTGCCT CAGGCTTGCACTGCTGACTC	119	58	NlaII	86 + 33
<i>MAGEA4</i>	NM_001011548.1	Forward Reverse	CACCAAGGAGAAGATCTGCCT CAGGCTTGCACTGCTGACTCT	118	58	NlaII	85 + 33
<i>PTPRC</i> (CD45)	NM_080923.2	Forward Reverse	TTAGGGACACGGCTGACTTC GCTTTGCCCTGTCAAAATA	106	62	NlaIII	72 + 34
<i>CD34</i>	NM_001773.2	Forward Reverse	AGGCCTCAGTGTCTACTGCTG TAGGATCCCCAGCTTTTCA	105	62	Tsp509I	64 + 41
<i>CD33</i>	NM_001772.3	Forward Reverse	GGCACCAACCTGACCTGT TCTTGTTCCTGAGCCATC	137	62	BsrI	105 + 32

<i>CD19</i>	NM_001770.5	Forward Reverse	CCTTGTGGGCATTCTTCATC CACGTTCCCGTACTGGTTCT	136	62	Hpy188III	97 + 39
<i>GYPC</i>	NM_002101.3	Forward Reverse	GAGAAGCCCCAACAGCAC AATCACACCTGCAATGACGA	187/130*	60	Hpy188III	114 + 73
<i>SCL4A1</i> (Band 3)	NM_000342.3	Forward Reverse	TCTTCAGGAACGTGGAGCTT CTGCTTTTCCTTGGAAGGTG	172	58	NotI	98 + 74
<i>HBA2</i>	NM_000517.4	Forward Reverse	GGCTCTGCCCAGGTTAAGG AGGCAGTGGCTTAGGAGCTT	164	60	HpyCH4IV	65 + 99
<i>TFRC</i> (CD71)	NM_001128148.1	Forward Reverse	ACAACAGCCAACTGCTTTCA TTCTCAGCATTCCCGAAATC	154	58	HaeIII	87 + 67
<i>KIT</i> (CD117)	NM_000222.2	Forward Reverse	CTTTAGGAAGCAGCCCTATC ATCTGCATCCAGCAAGTCT	140	60	Hpy188III	71 + 69
<i>EPOR</i>	NM_000121.3	Forward Reverse	CCAAGTTCGAGAGCAAAGC TCCTCGAGCTGGTAGGAGAA	157	58	EcoRI	117 + 40
<i>NFE2</i>	NM_006163.2	Forward Reverse	CTAGAGCCATCTGGGCTTTC ATGGCTCACTTGGAGCATTC	191	58	HaeIII	148 + 43
<i>KLF1</i>	NM_006563.3	Forward Reverse	GGATGACTTCCTCAAGTGGTG GTTGGTGAGGAGGAGATCCA	172	60	HaeIII	124 + 48
<i>ACTB</i>	NM_001101.3	Forward Reverse	GTGGGGCGCCCCAGGCACCA CTCCTTAATGTCACGCACGATTTC	540	58	n.v.	n.v.
<i>EEF1A1</i>	NM_001402.5	Forward Reverse	TGCCCCAGGACACAGAGACT CTGTGTCGGGGTTGTAGCCA	290	58	n.v.	n.v.
<i>GAPDH</i>	NM_002046.3	Forward Reverse	AATCCCATCACCATCTTCCAG GCCATCACGCCACAGTTTCC	380	58	n.v.	n.v.
<i>DHRS3</i>	NM_004753.4	Forward Reverse	GGGCATGAGAGTCAGGTTTC GCAGCCTGTGGAAGTATGC	159	59	n.v.	n.v.

\* durch die Primer werden beide Isoformen amplifiziert







**Abbildung 3.** PCR-basierte Amplifikation der einzelnen Fragmente und jeweiliger Restriktionsverdau. Die Größe des jeweiligen PCR Amplifikats und die Größe der „verdauten“ Fragmente befinden sich für jedes Transkript an der Seite, die Banden des DNA Größenmarkers sind auf der anderen Seite markiert. Die Probe für die Etablierung der jeweiligen Bedingungen ist unter jeder Abbildung genannt. In den meisten Fällen waren dies WTA Produkte einer Einzelzelle (EZ). Das beteiligte Restriktionsenzym ist oberhalb der jeweiligen Spalte genannt. „No RE“ bezeichnet Spalten mit „unverdauten“ Fragmenten, „NT“ bezeichnet die Leerkontrolle der PCR Reaktion. Panels, die durch zusammenfügen einzelner Spalten entstanden sind, wurden durch Sternchen gekennzeichnet und die zusammengefügte Teile mit weißen Linien getrennt.

## 2.8 Genspezifische PCR zur Detektierung ausgewählter Transkripte in Patientenproben

Alle WTAs der Einzelzellen wurden mittels PCR wie zuvor beschrieben getestet. Um die korrekte Amplifikation des Zieltranskripts nachzuweisen, wurden alle Proben, die ein positives Ergebnis aufwiesen, dem Restriktionsverdau zugeführt. Nur wenn nach dem Restriktionsverdau auch die entsprechenden Banden zu sehen waren, wurde die Probe als positiv für das entsprechende Zieltranskript gewertet.

## 2.9. Microarray basierte Transkriptomanalyse

Das Markieren und Hybridisieren der WTAs auf Genexpressions-Mikroarrays wurde gemäß veröffentlichter Protokolle (Hartmann et al., 2006) durchgeführt. Sowohl WTA Produkte als auch Referenz cDNA wurden in einer Reaktion mit 3% Formamid, 2,4 mM CP2-BGL Primer (TCAGAATTCATGCCGCCCCCGGCCC), dNTPs (0,35 mM dATP und dGTP, 0,3 mM dTTP

und dCTP) und 50 mM markierter Nukleotide versetzt. Die Markierung erfolgte für die WTA Produkte mit digoxigeniertem dUTP (Roche) und aminodigoxigeniertem dCTP (Perkin Elmer), für die Referenz cDNA mit biotinyliertem dUTP (Roche) und biotinyliertem dCTP (Invitrogen). Die Primer wurden anschließend mittels dem Enzym BGLI (30 U; Fermentas) entfernt und die markierten Proben mittels einer Säule (Qiagen PCR purification kit) gemäß den Herstellerangaben aufgereinigt.

Humane Oligonukleotidarrays (OpArray human Version 4; Operon) wurden für 45 min in einem Prähybridisierungspuffer (5 x SSC, 0,1% SDS, 0,1% BSA) bei 42°C inkubiert. Anschließend wurden die zuvor markierten Proben in einer Arraybooster Hybridisierungsstation (Implen) bei 42°C über Nacht auf den Arrays hybridisiert. Anschließend wurden die Array-Slides mehrfach bei 42°C gereinigt: zweimal mit 2 x SSC, 0,1% SDS für je 5 min, zweimal mit 0,5% SSC, 0,1% SDS für je 10 min und zweimal mit 0,1% SSC für 2,5 min. Unspezifische Bindungen wurden mittels 1%-iger Blocklösung für Nukleinsäurehybridisierung (Roche) abgesättigt. Dann wurden die Arrays mit 16 µg/ml anti-Dig-Cy5 und 18 µg/ml Streptavidin-Cy3 (beide Jackson Laboratories) unter Lichtschutz gefärbt. Überschüssiges Färbereagenz wurde durch zweimaliges Waschen (4x SSC, 0,2% Tween) entfernt und die Arrays in einem Genpix 4000A Scanner (Axon Instruments) ausgewertet.

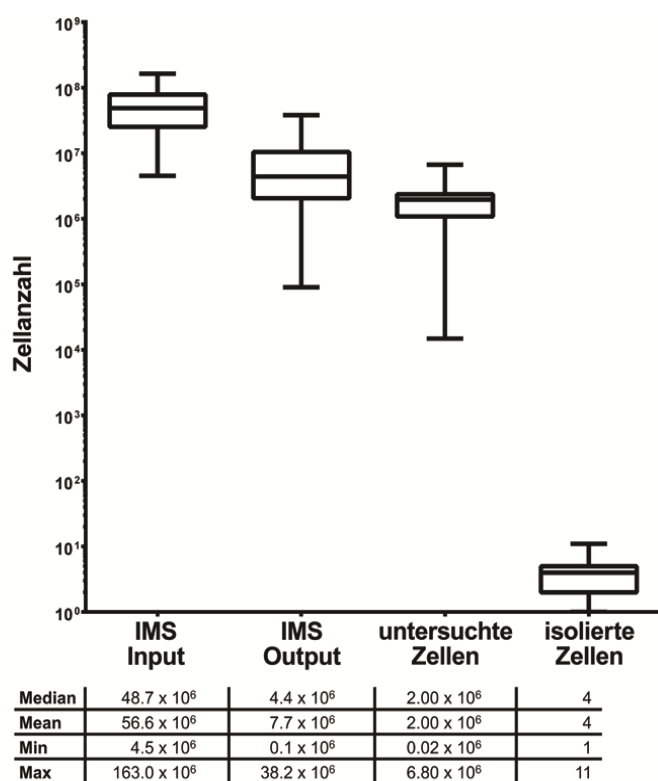
## **2.10 Statistische Auswertung**

Die statische Auswertung wurde mit SPSSv22 (IBM) und die hierarchische Clusteranalyse mit der Software Cluster durchgeführt (Eisen et al., 1998). Die bioinformatische Auswertung der gewonnenen Genexpressions-Mikroarray Daten wurde durch Herrn Matthias Maneck mit Hilfe der Software R-Bioconductor gemäß publizierter Algorithmen durchgeführt (VerMilyea et al., 2011).

### 3. Ergebnisse

#### 3.1 Nachweis von EpCAM<sup>+</sup> Zellen in den Knochenmarksproben und Korrelation mit klinischen Parametern

Das Vorhandensein von potentiellen DCCs im Knochenmark von Prostatakarzinompatienten ohne Metastasen (M0) wurde mittels EpCAM Färbung in 105 Knochenmarksaspiraten überprüft. Des Weiteren wurden Knochenmarksaspirate von zwei metastasierten Prostatakarzinompatienten (M1) und 18 männlichen Patienten ohne bekannte maligne Erkrankung (Kontrollgruppe) untersucht (Abb. 1). Um die Zielpopulation in den Patientenproben anzureichern, wurden zunächst CD45<sup>+</sup>, CD33<sup>+</sup>, CD11b<sup>+</sup> und Glycophorin A<sup>+</sup> Zellen entfernt, was zur durchschnittlichen Reduktion der Zellzahl um 86 % führte (Abb. 4). Danach wurden die Proben mittels direkter Immunfluoreszenz nach EpCAM<sup>+</sup> Zellen untersucht. Im Median wurden  $2 \times 10^6$  Zellen pro Knochenmarksprobe (Zellbereich von  $0,1 - 6,8 \times 10^6$  Zellen, Abb. 4) untersucht.



**Abbildung 4.** Übersicht Knochenmarksverarbeitung. Die *Box-Whisker Plots* enthalten die Zellanzahl der Proben vor bzw. nach der immunomagentischen Selektion (*IMS input* bzw. *output*), die Anzahl der untersuchten Zellen sowie die Anzahl der isolierten EpCAM<sup>+</sup> Zellen/Zellcluster. Die Zellanzahl auf der y-Achse ist in einer logarithmischen Skala dargestellt. Die Enden der Rechtecke stellen die 25. und 75. Perzentile dar, die Linie in der Mitte den Median. Die Whiskers entsprechen den jeweiligen Minimal- und Maximalwerten. Eine tabellarische Darstellung der genauen Werte befindet sich unter der Abbildung.

Vor Prostatektomie bzw. unfallchirurgischer OP konnten wir somit bei 65 (62%) der M0-Patienten und 10 (56%) der Kontrollpatienten EpCAM<sup>+</sup> Zellen im Knochenmark nachweisen. Die

Nachweisrate war nicht signifikant unterschiedlich ( $p = 0,79$ , zweiseitiger Fisher's Exakt Test; Tabelle 3). In beiden M1-Patienten wurden ebenfalls EpCAM<sup>+</sup> Zellen gefunden.

**Tabelle 2.** Häufigkeit von EpCAM<sup>+</sup> Zellen in den Knochenmarksproben aus PCa Patienten im M0- und M1-Stadium sowie Kontrollpatienten (N), Anzahl der isolierten EpCAM<sup>+</sup> Einzelzellen (EZ) und Zellclustern aus EpCAM<sup>+</sup> Zellen, sowie Anzahl der Zellen, die nach Qualitätskontrolle (QC) der WTA Produkte für die weiteren Analysen verwendet wurden.  $\Sigma$  – Summe.

	M0		N	M1	
untersuchte KM Proben	105		18	2	
EpCAM <sup>+</sup> Zellen detektiert in	65 (62%)		10 (56%)	2 (100%)	
Anzahl der isolierten EpCAM <sup>+</sup> Zellen	EZ	Cluster	EZ	EZ	Cluster
	222	5	51	12	4
Anzahl an Proben die nach Qualitätskontrolle für weitere Analysen verwendet wurden	119 (54%)	5 (100%)	28 (55%)	7 (58%)	3 (75%)
	$\Sigma$ : 55%		$\Sigma$ : 55%	$\Sigma$ : 62%	

Klinische Charakteristika der M0-Kohorte vor Prostatektomie werden in Tabelle 3 präsentiert. Wir konnten keinen signifikanten Zusammenhang zwischen dem Vorhandensein von EpCAM<sup>+</sup> Zellen im Knochenmark von M0-Prostatakarzinompatienten und einem der bekannten klinischen Parameter wie z.B. Gleason Score oder präoperativem PSA feststellen. Dies traf auch für weitere klinische Parameter wie Grading (G), Resektionsrand (R), Lymphgefäßinvasion (L), Gefäßinvasion (V) und perineurale Invasion (PnI) zu (Daten nicht gezeigt).

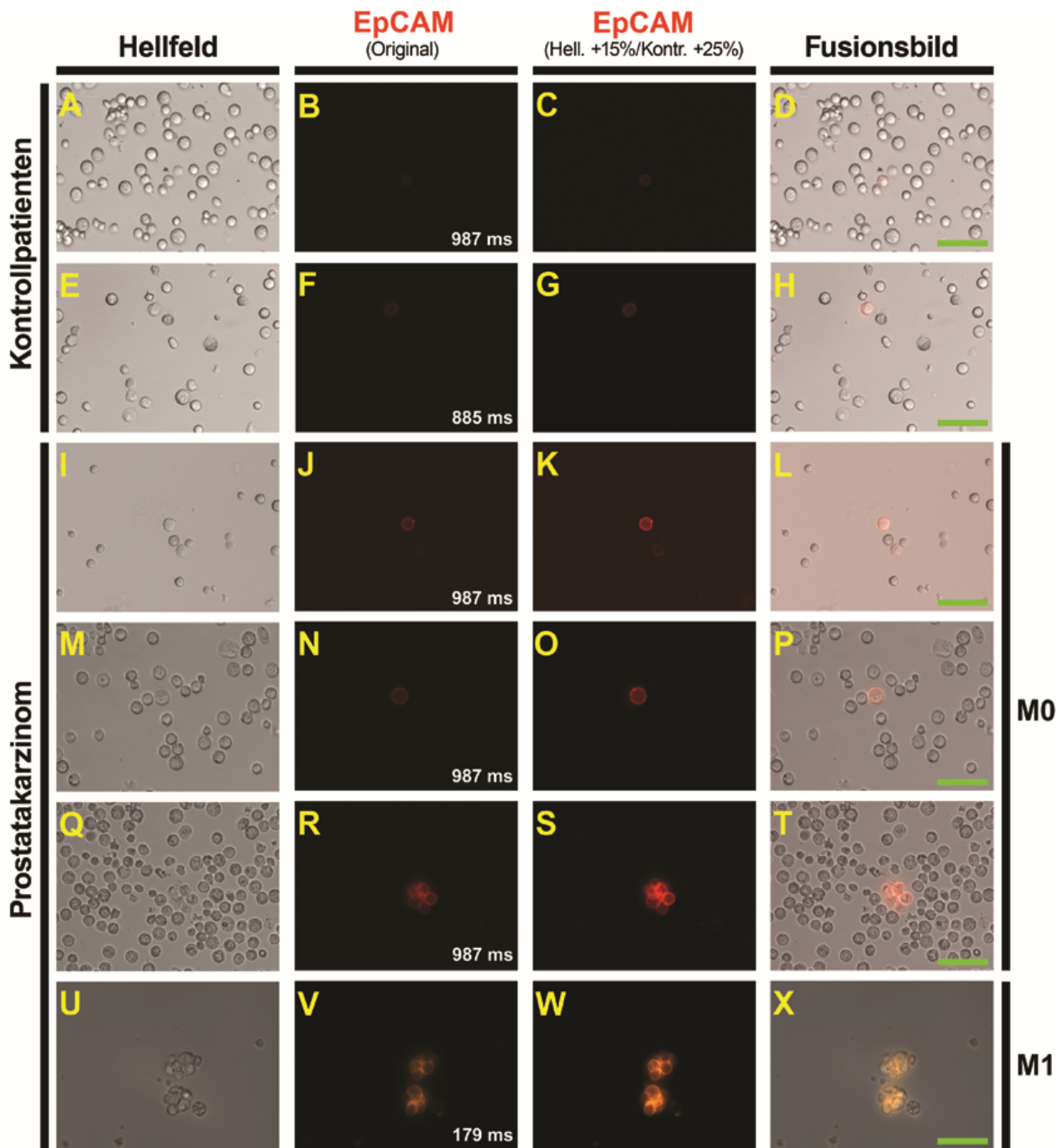
### 3.2 EpCAM Färbung in Knochenmarksproben von Prostatakarzinom-Patienten und der Kontrollgruppe

EpCAM<sup>+</sup> Zellen aus dem Knochenmark von Prostatakarzinompatienten (Abb. 5 I-X) wiesen meist eine höhere Färbeintensität auf als Zellen aus Kontrollproben (Abb. 5 A-H). Nur anhand der Färbeintensität, Größe oder Morphologie ist eine verlässliche Identifizierung von DCCs jedoch nicht möglich. Da das Phycoerythrin rasch ausbleicht, war es nicht möglich von jeder Zelle ein Photo für Intensitätsmessungen aufzunehmen. Ein Vergleich der vorhandenen Photos von EpCAM<sup>+</sup> Zellen aus Karzinompatienten und Kontrollpatienten mit EpCAM-gefärbten Prostataepithel-Zelllinien lässt dennoch den Schluss zu, dass EpCAM<sup>+</sup> Zellen aus dem Knochenmark von Patienten und Kontrollpersonen ähnliche Färbeeigenschaften aufweisen wie Zelllinien, die EpCAM in geringer Intensität exprimieren, wie z.B. PC3 oder DU145 (Abb. 6). Um diese Beobachtungen zu bestätigen, wurden aus einer Knochenmarkspalte eines Kontrollpatienten wie zuvor beschrieben CD45<sup>+</sup>, CD33<sup>+</sup>, CD11b<sup>+</sup> und GlyA<sup>+</sup> Zellen entfernt und die Probe danach geteilt. In eine Hälfte wurden anteilig 2% Zellen der schwach EpCAM-positiven DU145 Prostatakarzinom-Zelllinie

gegeben. Danach wurden die Teile erneut halbiert und mittels anti-EpCAM-PE beziehungsweise Isotyp-PE Antikörper gefärbt. In den Teilproben, die mit Isotyp-PE gefärbt wurden, konnte kein Fluoreszenzsignal detektiert werden (Abb. 7 A-H). Auf der anderen Seite wurden schwach EpCAM-positive Zellen in beiden Probenanteilen entdeckt, die mit anti-EpCAM-PE gefärbt wurden (Abb. 7 I-L). Wie erwartet, wiesen diese Zellen ein diskret schwächeres EpCAM-Signal auf im Vergleich zu DU145 Zellen (Abb. 7 M-P). Dennoch konnten nur 17% der zugemischten DU145 Population aufgrund der höheren EpCAM-Färbeintensität mittels durchflusszytometrischer Analyse identifiziert werden, was beweist, dass die Intensität der EpCAM Färbung kaum ausreicht um DCCs von anderen EpCAM<sup>+</sup> Knochenmarkszellen zu unterscheiden (Abb. 7 Q). Die durchflusszytometrische Analyse wurde von Frau Dr. Melanie Werner-Klein durchgeführt.

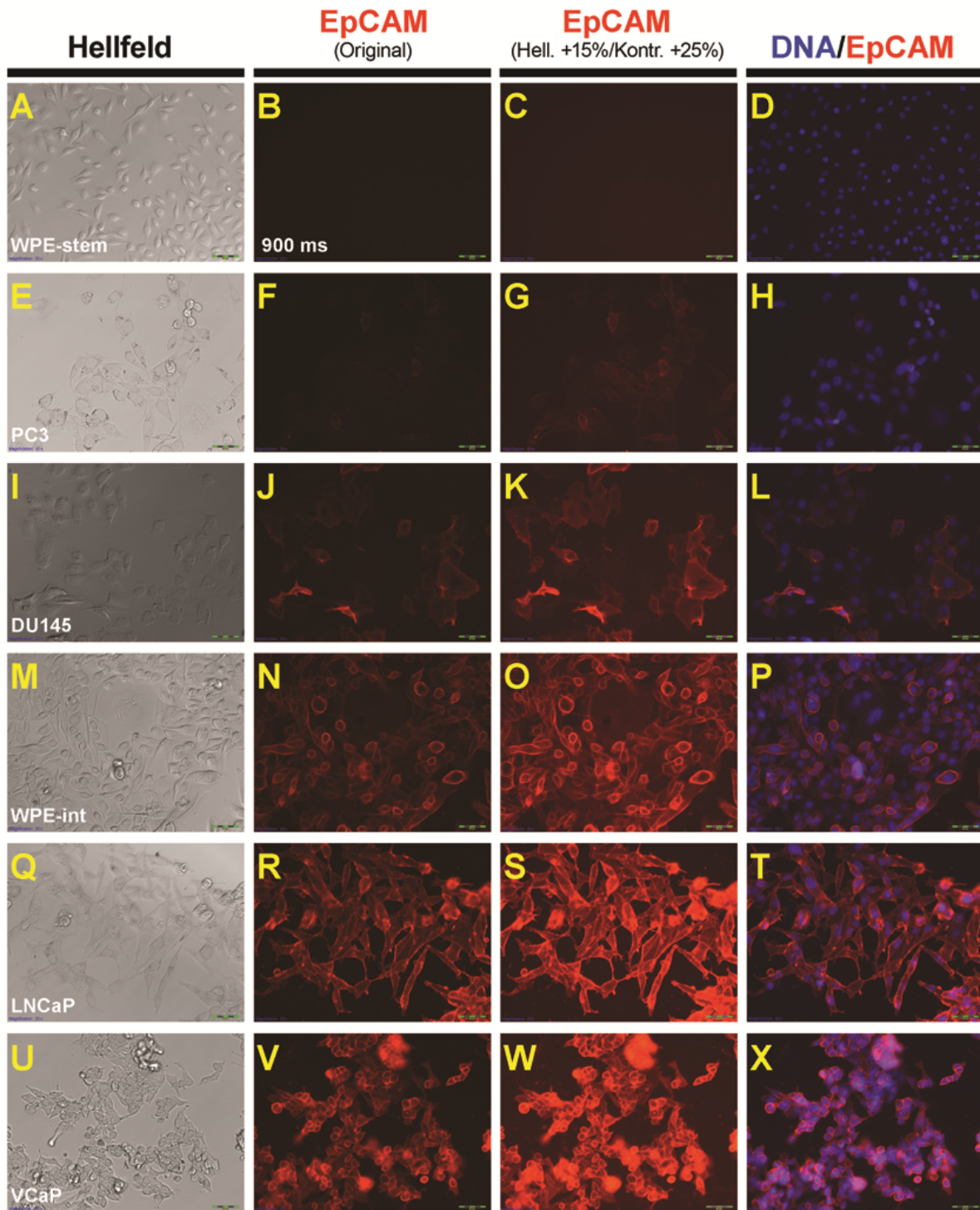
**Tabelle 3.** Klinische Parameter der M0 PCa Patienten mit und ohne EpCAM<sup>+</sup> Zellen im Knochenmark zum Zeitpunkt der Prostatektomie. Die Risikoeinteilung entspricht den Leitlinien der Europäischen Gesellschaft für Urologie (Heidenreich et al., 2012). Für die statistische Testung wurde der Chi-Quadrat Test verwendet. n.v. – nicht verfügbar

Klinische Parameter	Gesamt	EpCAM <sup>+</sup>		P
		ja	nein	
	n (%)	n (%)	n (%)	
M0 PCa Patienten	105	65 (62)	40 (38)	0,61
Alter				
≤ 60 Jahre	41 (39)	23 (56)	18 (44)	0,32
> 60 Jahre	64 (61)	42 (66)	22 (34)	
präoperatives PSA				
≤ 10 ng/μl	70 (67)	44 (63)	26 (37)	0,77
> 10 ng/μl	35 (33)	21 (60)	14 (40)	
Gleason Score				
2-6	40 (38)	24 (60)	16 (40)	0,95
7	51 (49)	32 (63)	19 (37)	
8-10	14 (13)	9 (64)	5 (36)	
T-Stadium				
T2	53 (51)	31 (59)	22 (41)	0,47
T3+4	52 (49)	34 (65)	18 (35)	
Lymphknotenstatus				
N0	58 (55)	33 (57)	25 (43)	0,32
N+	14 (13)	10 (71)	4 (29)	
n.v.	33			
Risikoeinteilung				
niedriges & mittleres Risiko	51 (49)	29 (57)	22 (43)	0,30
hohes & sehr hohes Risiko	54 (51)	36 (67)	18 (33)	



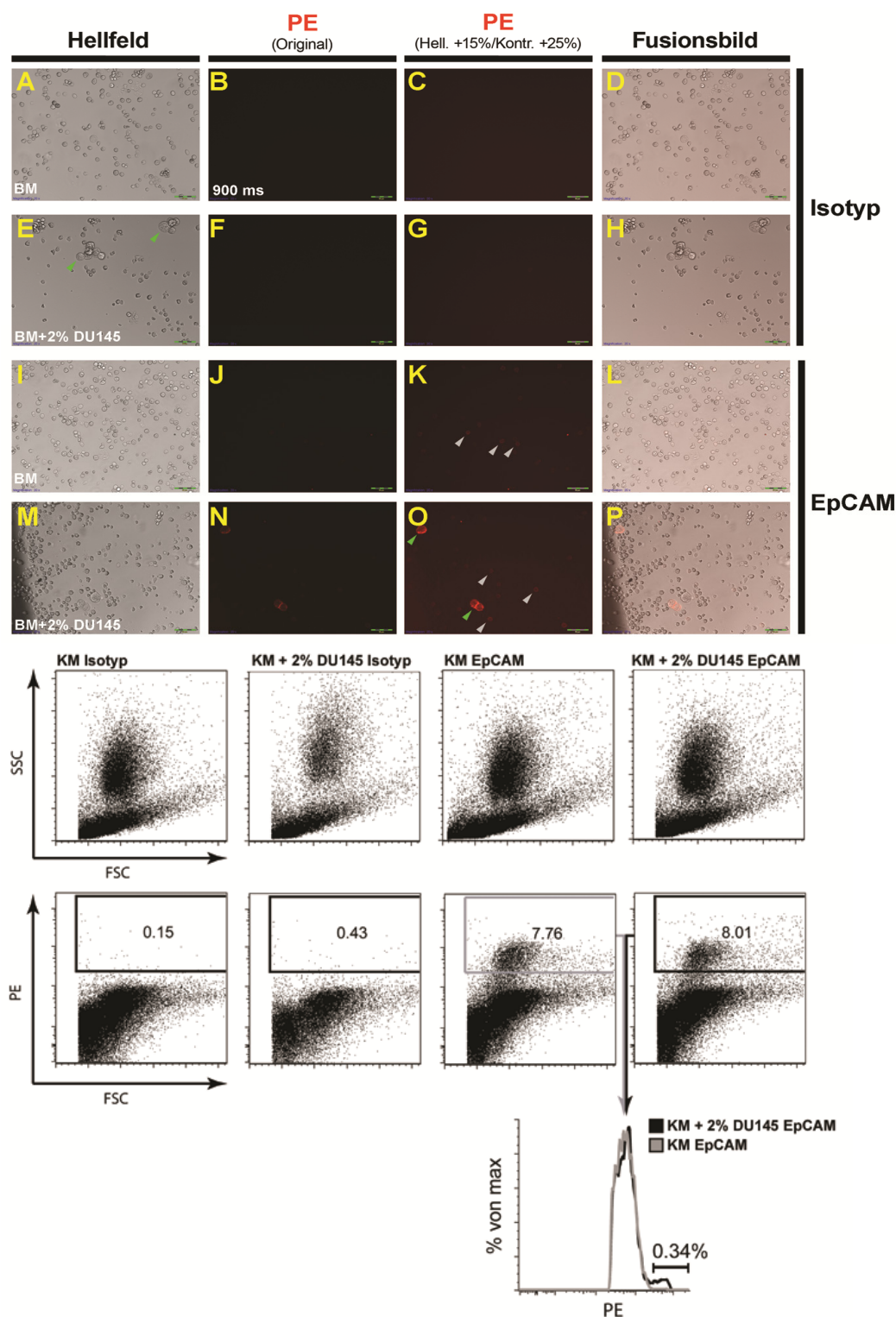
**Abbildung 5.** Intensität der EpCAM Färbung an den isolierten Zellen aus dem Knochenmark von PCa Patienten und Kontrollpatienten. Photos von EpCAM<sup>+</sup> Zellen aus dem Knochenmark von zwei Kontrollpatienten (**A-D** und **E-H**) und drei PCa Patienten, von denen zwei im M0-Stadium (**I-L** und **M-T**) und einer im M1-Stadium (**U-X**) war. Die Photos **A-D**, **E-H**, **I-L** und **M-P** zeigen EpCAM<sup>+</sup> Einzelzellen, die Photos **Q-T** und **U-X** zeigen EpCAM<sup>+</sup> Zellcluster. Die Bilder **A**, **E**, **I**, **M**, **Q**, und **U** repräsentieren lichtmikroskopische Photos. Die Bilder **B**, **C**, **F**, **G**, **J**, **K**, **N**, **O**, **R**, **S**, **V** und **W** repräsentieren Photos, die mithilfe des Fluoreszenzmikroskops (Cy3 Kanal) gemacht wurden. Die Zahlen in der unteren rechten Ecke in den Bildern **B**, **F**, **J**, **N**, **R**, und **V** ("EpCAM Original") repräsentieren die Belichtungsdauer. Die Photos **U-X** wurden mit einem anderen Mikroskop (Zeiss) geschossen als die Bilder **A-T** (Olympus). Die Bilder **C**, **G**, **K**, **O**, **S**, und **W** ("EpCAM (Hell./Kontr.)") wurden so bearbeitet, dass Helligkeit und Kontrast des Originalbildes erhöht wurden. Dies dient dem Zweck, die unterschiedliche Färbeintensität der EpCAM<sup>+</sup> Zellen in Tumor- und Kontrollpatienten besser sichtbar zu machen. Die Photos **D**, **H**, **L**, **P**, **T**, und **X** ergeben sich aus der Überlagerung des lichtmikroskopischen Bildes mit dem Fluoreszenzbild. Maßstab: 50 µm.





**Abbildung 6.** Photos der EpCAM Färbung an epithelialen Prostatakarzinomzelllinien. Der Name der jeweiligen Zelllinie befindet sich in der unteren linken Ecke des lichtmikroskopischen Bildes. Alle Bilder wurden bei einer Belichtungszeit von 900 ms aufgenommen. Bei den Photos der Spalte "EpCAM (Hell. +15%/Kontr. + 25%)" wurden Helligkeit und Kontrast wie angegeben verändert. Maßstab: 50 µm.



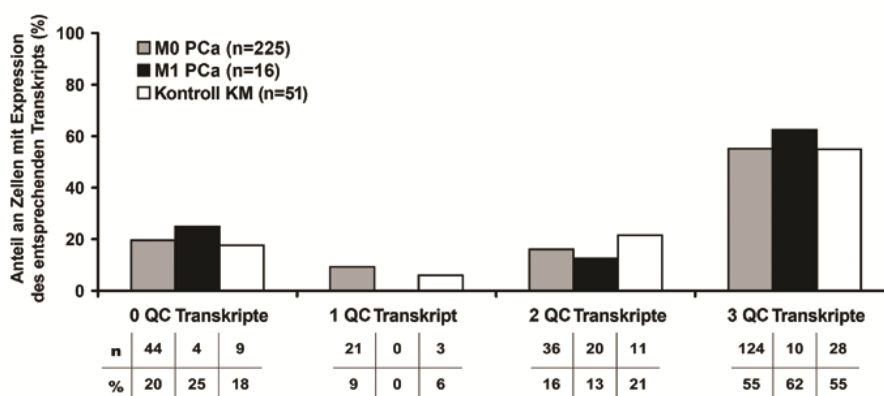


**Abbildung 7.** Photos und durchflusszytometrische Analyse der Knochenmarksprobe eines Kontrollpatienten. Die Probe wurde in 4 Teile aufgetrennt. In zwei Teile wurden anteilig 2% DU145 Zellen gemischt. Je eine Probe mit und eine ohne DU145 Zellen wurden mittels EpCAM-PE Antikörper gefärbt, die anderen beiden Proben mittels Isotyp-PE Antikörper. Die Isotyp-Färbung enthielt in keiner der beiden Proben (A-D und E-H; DU145 Zellen mit grünen Pfeilen markiert) gefärbte Zellen. Im Gegensatz dazu konnten in beiden Proben, die mittels EpCAM-PE Antikörper gefärbt wurden, mäßig EpCAM<sup>+</sup> Zellen (graue Pfeile) entdeckt werden (ohne DU145 I-L bzw. mit DU145 M-P; DU145 Zellen mit grünen Pfeilen markiert). Obwohl die Intensität der EpCAM Färbung in den DU145 Zellen leicht höher lag, konnten nur 17% der hinzugegebenen DU145-Population aufgrund der höheren Fluoreszenzwerte identifiziert werden (Q).

### 3.3 Isolation, Genom- und Transkriptomamplifikation von EpCAM<sup>+</sup> Einzelzellen

Im Mittel wurden 4 EpCAM<sup>+</sup> Zellen aus jeder positiven Knochenmarkspalte isoliert (siehe Abb. 4). Aus den insgesamt 65 EpCAM<sup>+</sup> Knochenmarkspalten von M0-Prostatakarzinompatienten wurden 220 EpCAM<sup>+</sup> Einzelzellen (Beispiele in Abb. 5, I-P) und 5 Cluster bestehend aus 2-5 untrennbaren Zellen (Beispiele in Abb. 5, Q-T) isoliert. Zwölf EpCAM<sup>+</sup> Einzelzellen und 4 Cluster bestehend aus 2-5 unteilbaren Zellen wurden aus den beiden M1-Prostatakarzinompatienten isoliert (Beispiele in Abb. 5, U-X). Des Weiteren wurden noch 51 EpCAM<sup>+</sup> Einzelzellen aus den Knochenmarkspalten von 10 Kontrollpatienten (Beispiele in Abb. 5, A-H, Tabelle 2) gesammelt. Im Knochenmark von Kontrollpatienten wurden keine EpCAM<sup>+</sup> Zellcluster gefunden. Nach Isolation der EpCAM<sup>+</sup> Einzelzellen wurde aus jeder Knochenmarkspalte ein Pool von ca. 2000-3000 Zellen isoliert.

Danach wurde von jeder Einzelzelle bzw. jedem Zellpool sowohl genomische DNA (gDNA) als auch mRNA isoliert. Die mRNA wurde der Transkriptom- (WTA), die gDNA der Genomamplifikation (WGA) zugeführt. Zur sinnvollen molekularen Analyse von Einzelzellen werden Proben von höchster Qualität benötigt. Deshalb wurden nur WTAs für die weiterführenden Untersuchungen verwendet, die ausgewählte Qualitätskriterien erfüllt haben (d.h. in allen Kontroll-PCRs für *ACTB*, *EEF1A1* und *GAPDH* positive Resultate erzielt haben, siehe auch „Material & Methoden“). Die drei Gruppen (M0, M1, Kontrollgruppe) unterschieden sich nicht bezüglich der Probenqualität ( $p = 0,75$ ; Pearson Chi-Quadrat Test; Tabelle 2 und Abb. 8). Nach WTA und Qualitätskontrolle wurden 119 EpCAM<sup>+</sup> Einzelzellen und 5 EpCAM<sup>+</sup> Zellcluster aus den M0-Patienten (55 % der hier isolierten Zellen) ausgewählt. Aus den Kontrollpatienten wählten wir 28 EpCAM<sup>+</sup> Einzelzellen (55 % der hier isolierten Zellen) und von den M1-Patienten 7 EpCAM<sup>+</sup> Einzelzellen und 3 EpCAM<sup>+</sup> Zellcluster (62 % der hier isolierten Zellen) aus (Tabelle 2, Abb. 8).

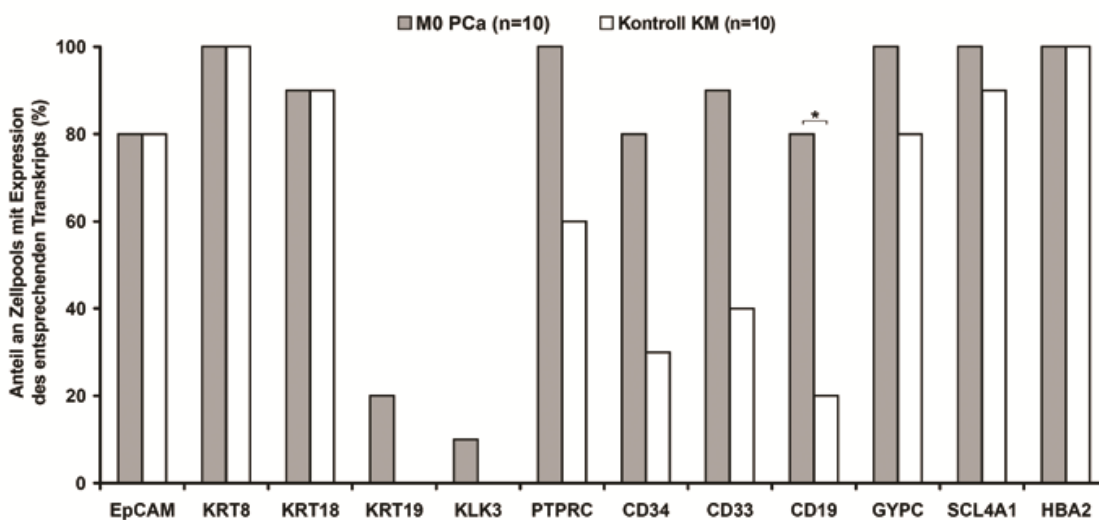


**Abbildung 8.** PCR Analyse zur Qualitätskontrolle der WTA Produkte aus jeder der untersuchten Gruppen. PCa Patienten im M0-Stadium (graue Balken) und im M1-Stadium (schwarze Balken). Kontrollpatienten (weiße Balken). Auf der x-Achse sind die Anzahl der vorhandenen QC-Transkripte (*ACTB*, *GAPDH*, *EEF1A1*) angegeben, die in den einzelnen Proben detektiert wurden. Die y-Achse stellt die Häufigkeit der Zellen in jeder Kategorie dar. *n* bezeichnet die Anzahl der untersuchten Einzelzellen. Genaue Zahlenangaben befinden sich unter dem Diagramm.

Des Weiteren wurden je 10 zufällig ausgewählte Zellpool-WTAs aus M0- und Kontrollpatienten sowie ein Pool eines M1-Patienten für die weiteren Untersuchungen verwendet.

### 3.4 Gezielte Transkriptomanalyse zur Identifikation einer DCC Signatur bei Patienten im M0 Stadium

Für die Suche nach einer DCC-spezifischen Signatur wurden zu Beginn ausgewählte Marker an den Zellpool-WTAs aus Prostatakarzinom- und Kontrollpatienten getestet. Diese Zellpools beinhalteten 2000 bis 3000 Zellen, die nach zuvor beschriebener Prozedur gewonnen wurden. Zwischen beiden Gruppen gab es keine signifikanten Unterschiede hinsichtlich der Häufigkeit von Proben, die epitheliale, prostata- oder tumorspezifische und erythrozytäre Transkripte exprimierten (Tabelle 5, Abb. 9 und Abb. 15 D-F). Zu unserer Überraschung war unter den hämatopoetischen Transkripten *CD19* signifikant häufiger in Zellpools von M0-Patienten verglichen mit Zellpools aus Kontrollpatienten exprimiert ( $p < 0,05$ ; Fisher's Exakt Test). Die Expression von *AR*, *ERG*, *CD44* und *PROM1* wurde nicht bei allen Zellpool-WTAs untersucht und deshalb nicht in die Analyse eingeschlossen.



**Abbildung 9.** Häufigkeit der Zellpools, die eines der untersuchten Transkripte exprimieren. Die jeweiligen Transkripte befinden sich auf der X-Achse. Die Ergebnisse für *KRT14*, *KRT6a*, *KRT5*, *MAGEA2* und *MAGEA4* wurden nicht dargestellt, da diese Transkripte in keiner der Proben detektiert wurden. Die genauen Häufigkeiten und Prozentangaben können Tabelle S3 entnommen werden. PCa Patienten im M0-Stadium (graue Balken) und im M1-Stadium (schwarze Balken). Kontrollpatienten (weiße Balken). Das Sternchen repräsentiert das Signifikanzniveau  $p < 0,05$ .  $n$  entspricht der Anzahl an untersuchten Zellpools.

Da sich in der Untersuchung der Zellpools aus den Knochenmarksproben kein Marker spezifisch für Tumor- oder Kontrollpatienten erwies, wurde sich auf die gezielte Transkriptomanalyse der EpCAM<sup>+</sup> Einzelzellen konzentriert. Zunächst überprüfen wir die Genexpression von *EPCAM* und verschiedenen Zytokeratinen. Die Anzahl der EpCAM<sup>+</sup> Einzelzellen, die das *EPCAM* Transkript exprimierten, war unter den Zellen aus M0-Prostatakarzinom-Patienten (88%) signifikant höher als

unter den Zellen aus dem Knochenmark von Kontrollpatienten (50%;  $p < 0,001$ ; Pearson Chi-Quadrat Test; Tabelle 5, Abb. 10A und Abb. 13).

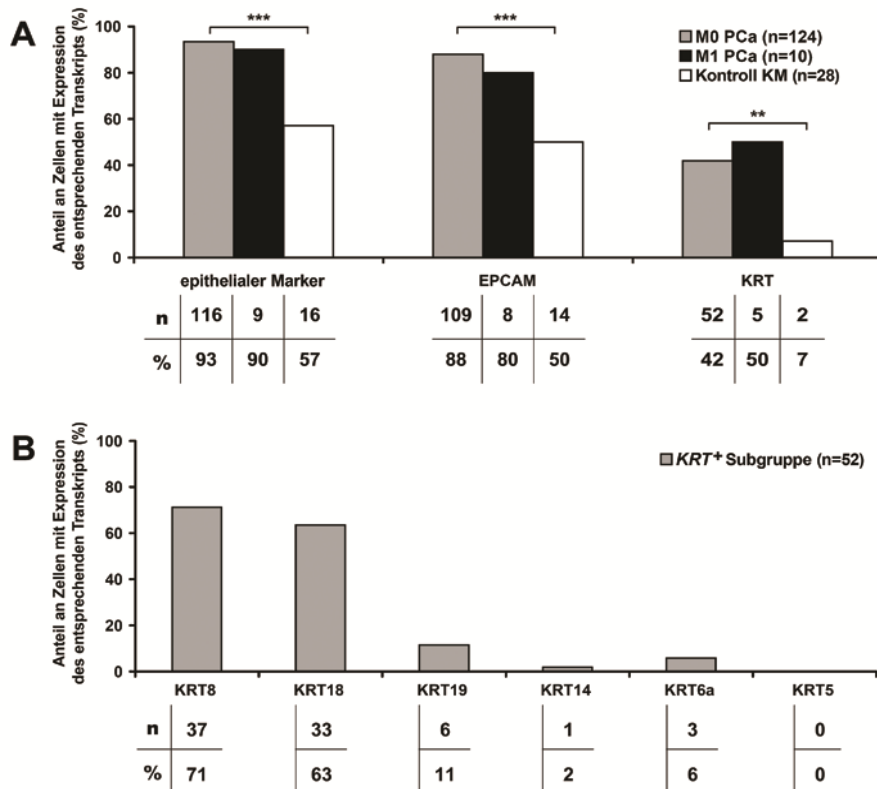
Zweiundvierzig Prozent der Einzelzellen aus M0-Patienten exprimierten mindestens ein Zytokeratin (Anhang Abb. 10A), was sich signifikant von den Einzelzellen der Kontrollgruppe unterschied, in der lediglich 7% der Einzelzellen ein Zytokeratintranskript exprimierten ( $p < 0,01$ ; Pearson Chi-Quadrat Test). Der Unterschied zwischen den Zellen aus Tumor- und Kontrollpatienten war sogar noch ausgeprägter, wenn man diejenigen Zellen verglich, die sowohl *KRT* als auch *EpCAM* Expression aufwiesen. Zellen aus dem Knochenmark von M0-Prostatakarzinompatienten koexprimierten signifikant häufiger *EPCAM* und *KRT* Transkripte verglichen mit Zellen aus der Kontrollgruppe ( $p < 0,002$ ; Fisher's Exakt Test, Tabelle 4). In der Gruppe der *KRT*-exprimierenden Einzelzellen aus M0-Patienten wurden *KRT8* und *KRT18* am häufigsten detektiert (Tabelle 5, Abb. 10B und Abb. 13). Obwohl sie insgesamt in niedriger Häufigkeit exprimiert wurden, waren *KRT8* und *KRT18* signifikant häufiger in Einzelzellen aus M0-Patienten zu finden als bei Kontrollpatienten (*KRT8* – 30% vs. 4%,  $p < 0,01$ ; *KRT18* – 27% vs. 7%,  $p < 0,05$ ; beides Pearson Chi-Quadrat Test; Tabelle 5). Transkripte von *KRT19*, *KRT14* und *KRT6* wurden in weniger als 5% der Einzelzellen aus M0-Patienten und in keiner der Einzelzellen aus der Kontrollgruppe detektiert. Transkripte von *KRT5* wurden in keiner der beiden Gruppen gefunden (Tabelle 5, Abb. 10B und Abb. 13). In der Gruppe der Einzelzellen aus M0-Patienten exprimierten 93% der Einzelzellen mindestens eines der getesteten epithelialen Transkripte. Im Vergleich hierzu war dies nur bei 57% der Einzellzellen aus Kontrollpatienten der Fall ( $p < 0,001$ ; Fisher's Exakt Test). Diese Verteilung zwischen beiden Gruppen entspricht in etwa der Verteilung der *EpCAM* Expression (Abb. 10A).

**Tabelle 4.** Verteilung der *EpCAM*<sup>+</sup> Einzelzellen aus dem Knochenmark von M0- (a) und M1-PCa (c) Patienten sowie (b) Kontrollpatienten (N), die *EPCAM* und *KRT* Transkripte koexprimieren.  $\Sigma$  – Summe.

(a) M0	<i>KRT</i> <sup>+</sup>	<i>KRT</i> <sup>−</sup>	$\Sigma$	(b) N	<i>KRT</i> <sup>+</sup>	<i>KRT</i> <sup>−</sup>	$\Sigma$
<i>EPCAM</i> <sup>+</sup>	45 (36%) <sup>a</sup>	64 (52%) <sup>b</sup>	109 (88%)	<i>EPCAM</i> <sup>+</sup>	0 (0%) <sup>c</sup>	14 (50%) <sup>d</sup>	14 (50%)
<i>EPCAM</i> <sup>−</sup>	7 (6%)	8 (6%)	15 (12%)	<i>EPCAM</i> <sup>−</sup>	2 (7%)	12 (43%)	14 (50%)
$\Sigma$	52 (42%)	72 (58%)	124	$\Sigma$	2 (7%)	26 (93%)	28
(c) M1	<i>KRT</i> <sup>+</sup>	<i>KRT</i> <sup>−</sup>	$\Sigma$				
<i>EPCAM</i> <sup>+</sup>	4 (40%)	4 (40%)	8 (80%)				
<i>EPCAM</i> <sup>−</sup>	1 (10%)	1 (10%)	2 (20%)				
$\Sigma$	5 (50%)	5 (50%)	10				

Bemerkung:

a,b vs. c,d –  $p < 0.002$ ; Fisher's Exakt Test



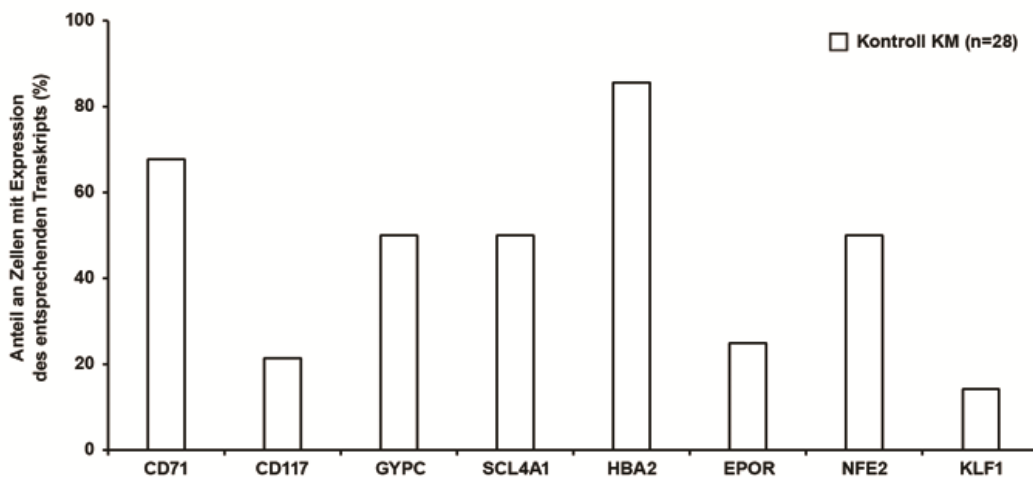
**Abbildung 10.** Überblick über die Expression epithelialer Transkripte. **A.** Häufigkeit der Zellen, die eines der untersuchten Transkripte exprimieren, sortiert nach untersuchten Gruppen. PCa Patienten im M0-Stadium (graue Balken) und im M1-Stadium (schwarze Balken). Kontrollpatienten (weiße Balken). Die jeweiligen Transkripte befinden sich auf der x-Achse. Die Bezeichnung epithelialer Marker beinhaltet die Expression mindestens eines der folgenden Transkripte: *EPCAM*, *KRT8*, *KRT18*, *KRT19*, *KRT14*, *KRT6a*, und *KRT5*. KRT bezeichnet die Expression mindestens eines Zytokeratin-Transkripts (*KRT8*, *KRT18*, *KRT19*, *KRT14*, *KRT6a*, und *KRT5*). Die Sternchen repräsentieren das Signifikanzniveau: \*\* –  $p < 0.01$ , \*\*\* –  $p < 0.001$ ;  $n$  entspricht der Anzahl an untersuchten Einzelzellen. Genaue Zahlenangaben befinden sich unter dem Diagramm. **B.** Häufigkeit der Zellen aus M0-PCa Patienten, die eines der getesteten Zytokeratin-Transkripte exprimieren. Die Häufigkeit bezieht sich hier auf diejenigen Zellen, die mindestens eines der getesteten Zytokeratin-Transkripte exprimieren.  $n$  entspricht der Anzahl an untersuchten Einzelzellen. Genaue Zahlenangaben befinden sich unter dem Diagramm.

Als nächsten Schritt haben wir das Vorhandensein prostata- und tumorspezifischer Transkripte überprüft. *KLK3* (PSA) Transkripte wurden lediglich in Einzelzellen eines M0-Prostatakarzinom-Patienten gefunden, dessen Zellpool ebenfalls *KLK3* exprimierte. Insgesamt konnte somit *KLK3* in 9 (7%) der Einzelzellen aus M0-Patienten und in keiner der Zellen der Kontrollgruppe detektiert werden (nicht signifikant, Fisher's Exakt Test, Tabelle 5 und Abb. 13). Transkripte für AR wurden in 6 (5%) der Einzelzellen aus M0-Patienten und in einer der Zellen (4%) aus der Kontrollgruppe identifiziert (nicht signifikant, Fisher's Exakt Test, Tabelle 5 und Abb. 13). In 17 Einzelzellen (14 %) aus dem Knochenmark von M0 Patienten, aber keiner der Zellen aus der Kontrollgruppe konnten *ERG* Transkripte nachgewiesen werden ( $p < 0,05$ ; Fisher's Exakt Test). Die Expression von *CD44* und *PROM1* (CD133) – zwei Marker, die als charakteristisch für Prostatastammzellen beschrieben wurden – konnte in 60 (48 %) bzw. 6 (5%) der Einzelzellen von M0-Patienten gefunden werden. Eine parallele Expression beider Marker wiesen nur 4 von 124 EpCAM<sup>+</sup> Einzelzellen aus M0 Patienten auf (3% der Zellen). Der Anteil von Einzelzellen mit *CD44* Expression war signifikant

höher in M0-Patienten verglichen mit der Kontrollgruppe (4% mit *CD44* Expression,  $p < 0,001$ ; Fisher's Exakt Test, Tabelle 5 und Abb. 13). Die Expression von *PROM1* unterschied sich nicht signifikant. *MAGEA2* und *MAGEA4* konnten in keiner der untersuchten EpCAM<sup>+</sup> Einzelzellen gefunden werden.

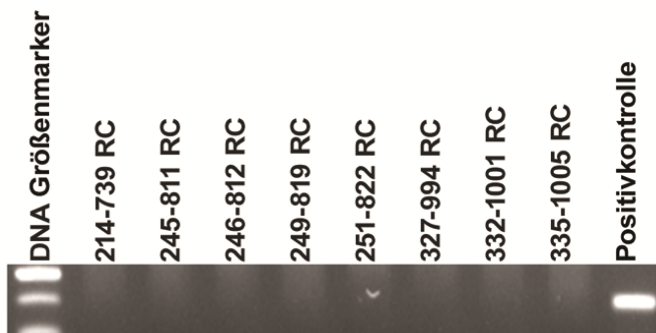
In Anbetracht der niedrigen Expressionshäufigkeit der histogenetischen oder tumorassoziierten Transkripte wollten wir untersuchen ob hämatopoetische Transkripte hilfreich dabei wären, EpCAM<sup>+</sup> Zellen aus Tumorpatienten gegenüber EpCAM<sup>+</sup> Zellen aus Kontrollpatienten abzugrenzen. Deshalb bestimmten wir den Expressionsstatus ausgewählter hämatopoetischer Transkripte in den isolierten Einzelzellen. Zu unserer Überraschung wurde unter den Einzelzellen aus M0-Patienten *PTPRC* (CD45) am häufigsten exprimiert (19% der Einzelzellen; Tabelle 5 und Abb. 13). Insgesamt exprimierten 23% der EpCAM<sup>+</sup> Einzelzellen aus M0-Prostatakarzinompatienten mindestens eines der ausgewählten hämatopoetischen Transkripte (*PTPRC*, *CD34*, *CD33* oder *CD19*). In den analog dazu untersuchten Einzelzellen der Kontrollgruppe exprimierten insgesamt 29% der EpCAM<sup>+</sup> Einzelzellen mindestens eines der ausgewählten hämatopoetischen Transkripte, wobei wiederum *PTPRC* am häufigsten war (21% der Zellen; Tabelle 5 und Abb. 13). Die Expressionshäufigkeit der hämatopoetischen Transkripte unterschied sich zwischen beiden Gruppen nicht signifikant ( $p = 0,5$ ; Pearson Chi-Quadrat Test).

Darüber hinaus haben wir die Expression ausgewählter Transkripte analysiert, die als spezifisch für die erythrozytäre Linie gelten. Hierzu wurde zunächst die Expression von *TFRC* (CD71), *KIT* (CD117), *GYPC* (CD236, Glycophorin C), *SCL4A1* (CD33, Bande 3 Protein), *HBA2* (Hämoglobin  $\alpha$  2), *EPOR*, *NFE2* und *KLF1* in den EpCAM<sup>+</sup> Einzelzellen der Kontrollgruppe untersucht. Die am häufigsten exprimierten Transkripte waren hierbei *HBA2* (86% der Zellen) und *TFRC* (68%), gefolgt von *GYPC*, *SCL4A1* und *NFE2* (je 50% der Einzelzellen; siehe Tabelle 5 und Abb. 11). Insgesamt exprimierten 27 der 28 untersuchten Kontrollzellen mindestens einen der erythrozytären Marker (96%). Da jedoch das Vorhandensein von *TFRC* und *KIT* auch in Prostataepithelzellen beschrieben wurde (Liu et al., 2002) und einige Marker eine vergleichbar geringe Häufigkeit in der Kontrollgruppe aufwiesen, wurden lediglich *HBA2*, *GYPC* und *SCL4A1* auf deren Expression in den EpCAM<sup>+</sup> Einzelzellen aus M0 Prostatakarzinompatienten untersucht. Überraschenderweise konnten auch hier keine signifikanten Unterschiede zwischen den beiden Gruppen detektiert werden (Pearson Chi-Quadrat Test, Tabelle 5 und Abb. 13). Auch in den M0 Patienten exprimierten 96% der Einzelzellen mindestens eines der untersuchten erythrozytären Transkripte.



**Abbildung 11.** Häufigkeit der EpCAM<sup>+</sup> Einzelzellen aus den Kontrollpatienten, die ausgewählte erythrozytäre Transkripte exprimieren. Die jeweiligen Transkripte befinden sich auf der x-Achse. Die genauen Häufigkeiten und Prozentangaben können Tabelle S3 entnommen werden.

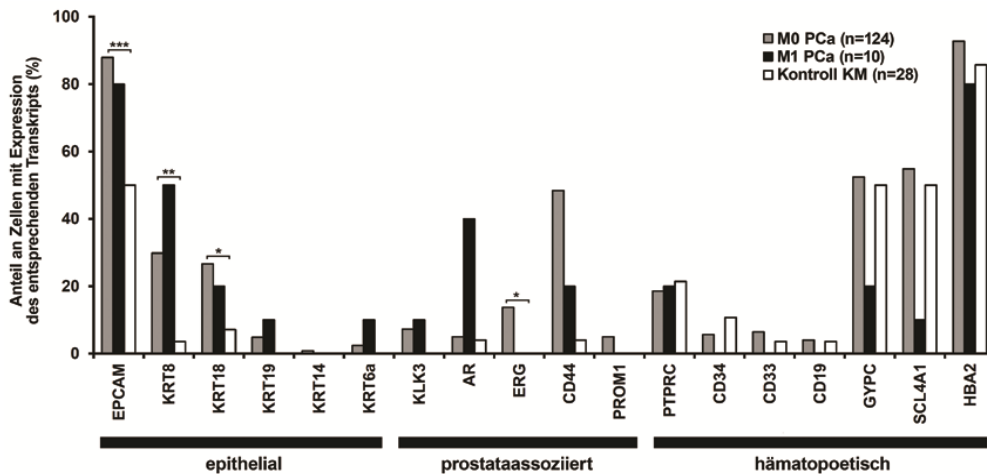
Diese unerwarteten Beobachtungen brachten die Frage auf, ob die WTAs der Einzelzellen eventuell durch extrazelluläre erythrozytäre Transkripte verunreinigt wurden. Dies könnte z.B. durch interzelluläre mRNA Übertragung oder Freisetzung der mRNA durch Zellzerstörung während der Probenverarbeitung geschehen. Deshalb wurden die WTA-Negativkontrollen (siehe auch Material und Methoden) von 40 Patienten nach dem Vorhandensein von *HBA2* Transkripten untersucht. Da alle dieser getesteten Proben negativ waren (Abb. 12) kann eine Verunreinigung der Einzelzell-WTA Produkte sicher ausgeschlossen werden.



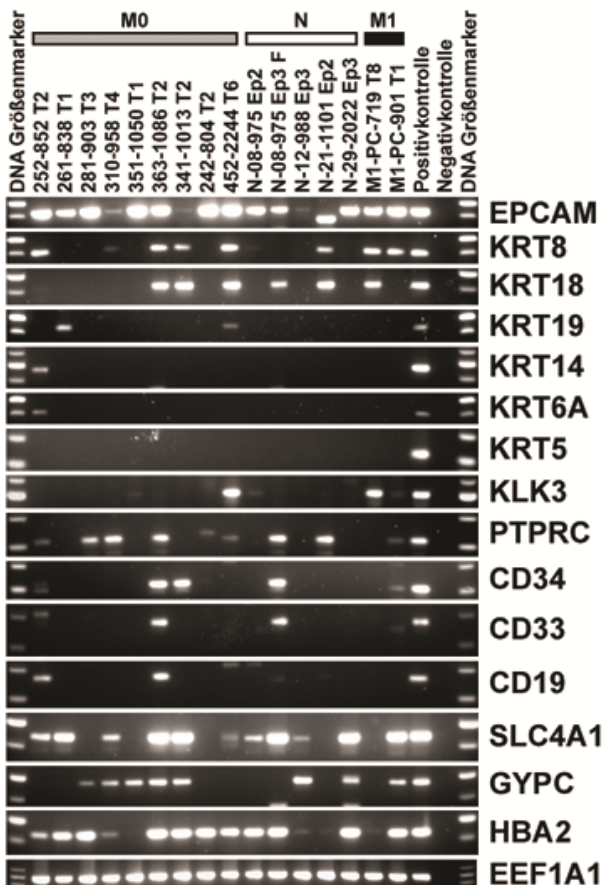
**Abbildung 12.** Untersuchung der WTA Kontrollen auf das Vorhandensein des *HBA2* Transkripts. Insgesamt wurden 40 Proben analysiert, die alle ein negatives Resultat aufwiesen. Hier sind exemplarisch 8 Proben gezeigt.

In Zusammenschau aller analysierter Marker zeigten sich durch die Koexpression von epithelialen, tumorassoziierten und erythrozytären Transkripten in EpCAM<sup>+</sup> Einzelzellen aus Tumor- und Kontrollpatienten unerwartete Expressionsmuster (Abb. 14 und Abb. 15A-B).





**Abbildung 13.** Häufigkeit der Zellen, die eines der untersuchten Transkripte exprimieren. Die jeweiligen Transkripte befinden sich auf der x-Achse. Die Ergebnisse für *KRT5*, *MAGEA2* und *MAGEA4* wurden nicht dargestellt, da diese Transkripte in keiner der Proben detektiert wurden. Die genauen Häufigkeiten und Prozentangaben können Tabelle S3 entnommen werden. PCa Patienten im M0-Stadium (graue Balken) und im M1-Stadium (schwarze Balken), Kontrollpatienten (weiße Balken). Die Sternchen repräsentieren das Signifikanzniveau: \* –  $p < 0.05$ , \*\* –  $p < 0.01$ , \*\*\* –  $p < 0.001$ ;  $n$  entspricht der Anzahl an untersuchten Einzelzellen.



**Abbildung 14.** PCR basierte Transkriptomanalyse der WTA aus ausgewählten EpCAM<sup>+</sup> Einzelzellen bzw. EpCAM<sup>+</sup> Zellclustern aus PCa Patienten und Kontrollpatienten. Die Photos eines jeden Gels wurden zurechtgeschnitten, sodass sie die jeweiligen Zielfragmente und die angrenzenden Banden des Größenmarkers enthielten. Die Größe der jeweiligen Fragmente kann Tabelle S1 und Abb. S2 entnommen werden. Die Probennummer ist über den Spalten abgebildet, das untersuchte Transkript neben dem jeweiligen Gelphoto. Helligkeit und Kontrast wurden mittels „Auto Kontrast“ Tool des Programms Adobe Photoshop angepasst.



**Tabelle 5.** Anzahl und Häufigkeit der Proben aus dem Knochenmark von M0- und M1-PCa Patienten und Kontrollpatienten, die die jeweiligen Transkripte exprimieren. Der Zellpool eines M1-Patienten, der ebenfalls untersucht wurde, ist hier aus Übersichtsgründen nicht abgebildet. Das Ergebnis kann Abb. S7 entnommen werden. Für statistische Tests wurden nur die Ergebnisse der Expressionsanalysen aus Einzelzellen von M0-PCa und Kontrollpatienten, sowie deren Zellpools verwendet. Signifikante Unterschiede wurden mit Fußnote gekennzeichnet (Pearsons Chi-Quadrat bzw. Fisher's Exakt Test). n.t. – nicht getestet, *KRT5*, *MAGEA2* und *MAGEA4* nicht dargestellt.

	EpCAM <sup>+</sup> Einzelzellen M0 PCa n (%)	EpCAM <sup>+</sup> Einzelzellen M1 PCa n (%)	EpCAM <sup>+</sup> Einzelzellen Kontrollpatienten n (%)	Zellpools M0 PCa n (%)	Zellpools Kontrollpatienten n (%)
Anzahl der Proben	124	10	28	10	10
<i>EPCAM</i>	109 (88) <sup>a</sup>	8 (80)	14 (50) <sup>a</sup>	8 (80)	8 (80)
<i>KRT8</i>	37 (30) <sup>b</sup>	5 (50)	1 (4) <sup>b</sup>	10 (100)	10 (100)
<i>KRT18</i>	33 (27) <sup>c</sup>	2 (20)	2 (7) <sup>c</sup>	9 (90)	9 (90)
<i>KRT19</i>	6 (5)	1 (10)	0 (0)	2 (20)	0 (0)
<i>KRT14</i>	1 (1)	0 (0)	0 (0)	0 (0)	0 (0)
<i>KRT6a</i>	3 (2)	1 (10)	0 (0)	0 (0)	0 (0)
<i>KLK3</i>	9 (7)	1 (10)	0 (0)	1 (10)	0 (0)
<i>AR</i>	6 (5)	4 (40)	1 (4)	n.t.	n.t.
<i>ERG</i>	17 (14) <sup>d</sup>	0 (0)	0 (0)	n.t.	n.t.
<i>CD44</i>	60 (48) <sup>a</sup>	2 (20)	1 (4) <sup>a</sup>	n.t.	n.t.
<i>PROM1</i>	6 (5)	0 (0)	0 (0)	n.t.	n.t.
<i>PTPRC</i>	23 (19)	2 (20)	6 (21)	10 (100)	6 (60)
<i>CD34</i>	7 (6)	0 (0)	3 (11)	8 (80)	3 (30)
<i>CD33</i>	8 (7)	0 (0)	1 (4)	9 (90)	4 (40)
<i>CD19</i>	5 (4)	0 (0)	2 (7)	8 (80) <sup>d</sup>	2 (20) <sup>d</sup>
<i>TFRC</i>	n.t.	n.t.	19 (68)	n.t.	n.t.
<i>KIT</i>	n.t.	n.t.	6 (21)	n.t.	n.t.
<i>EPOR</i>	n.t.	n.t.	7 (25)	n.t.	n.t.
<i>NFE2</i>	n.t.	n.t.	14 (50)	n.t.	n.t.
<i>KLF1</i>	n.t.	n.t.	4 (14)	n.t.	n.t.
<i>GYPC</i>	65 (52)	2 (20)	14 (50)	10 (100)	8 (80)
<i>SCL4A1</i>	68 (55)	1 (10)	14 (50)	10 (100)	9 (90)
<i>HBA2</i>	115 (93)	8 (80)	24 (86)	10 (100)	10 (100)

Bemerkung: <sup>a</sup> p<0.001, Pearson Chi- Quadrat <sup>b</sup> p<0.01, Pearson Chi- Quadrat  
<sup>c</sup> p<0.05, Pearson Chi-Quadrat <sup>d</sup> p<0.05, Fisher's Exakt Test

### **3.5 PCR Profiling von EpCAM<sup>+</sup> Zellen aus dem Knochenmark von metastasierten Patienten**

Für den Großteil der analysierten Gene ergab sich eine ähnliche Verteilung der Transkript-positiven Zellen unter den EpCAM<sup>+</sup> Einzelzellen und EpCAM<sup>+</sup> Zellclustern aus M1-Prostatakarzinompatienten im Vergleich zu den Einzelzellen aus den M0-Patienten (Tabelle 5, Abb. 10A, Abb. 13, Abb. 14 und Abb. 15C). Da wir nur zwei Knochenmarksproben und 7 EpCAM<sup>+</sup> Einzelzellen sowie 3 EpCAM<sup>+</sup> Zellcluster von M1-Patienten zur Verfügung hatten, wurden keine weiteren statistischen Tests vergleichend zu anderen Gruppen durchgeführt.

### **3.6 Genexpressionsprofile und phänotypische Heterogenität**

Keines der getesteten Transkripte konnte verlässlich zwischen EpCAM<sup>+</sup> Zellen aus Tumor- und Kontrollpatienten unterscheiden. Deshalb wurde untersucht, ob eine Kombination der getesteten Marker dabei hilfreich wäre, DCCs korrekt zu identifizieren. Nach hierarchischer Clusteranalyse zeigten sich alle EpCAM<sup>+</sup> Einzelzellen und Zellcluster (aus M0-, M1- und Kontrollpatienten) gleichmäßig zwischen den Dendrogrammgruppen verteilt (Abb. 16A). Die Analyse der Zellpools erbrachte zwei Hauptgruppen; innerhalb dieser Cluster waren jedoch Pools aus Tumor- und Kontrollpatienten gleichmäßig verteilt (Abb. 16B).

Von 35 (54%) der EpCAM<sup>+</sup> M0-Patienten konnten zwei oder mehr Zellen isoliert werden. Unter diesen 35 Patienten gruppieren sich die Zellen aus 14 Patienten (40%) zusammen. Jedoch besaßen 12 von 14 Patienten mindestens eine Zelle, die sich nicht mit den anderen grupperte, wohingegen sich in lediglich zwei Fällen alle Zellen in einem Cluster befanden. Die Mehrzahl aller Proben, aus denen mindestens zwei Zellen isoliert wurden, beinhalteten somit Zellen, die eine Heterogenität bezüglich der analysierten Transkripte aufwiesen. Im Fall der Kontrollpatienten wiesen fünf von zehn Proben mindestens zwei EpCAM<sup>+</sup> Zellen auf. Auch hier gruppieren lediglich Zellen von zwei der fünf Patienten (40%) miteinander. Ebenso gab es je mindestens eine Zelle, die sich nicht mit den anderen grupperte. Der Unterschied zwischen beiden Gruppen war nicht signifikant ( $p = 1$ ; Fisher's Exakt Test). Die Zellen eines M1-Patienten waren hauptsächlich in einem gemeinsamen Cluster zu finden, jedoch gab es auch hier mehrere Zellen, die sich zu anderen Clustern grupperten (Abb. 16A).

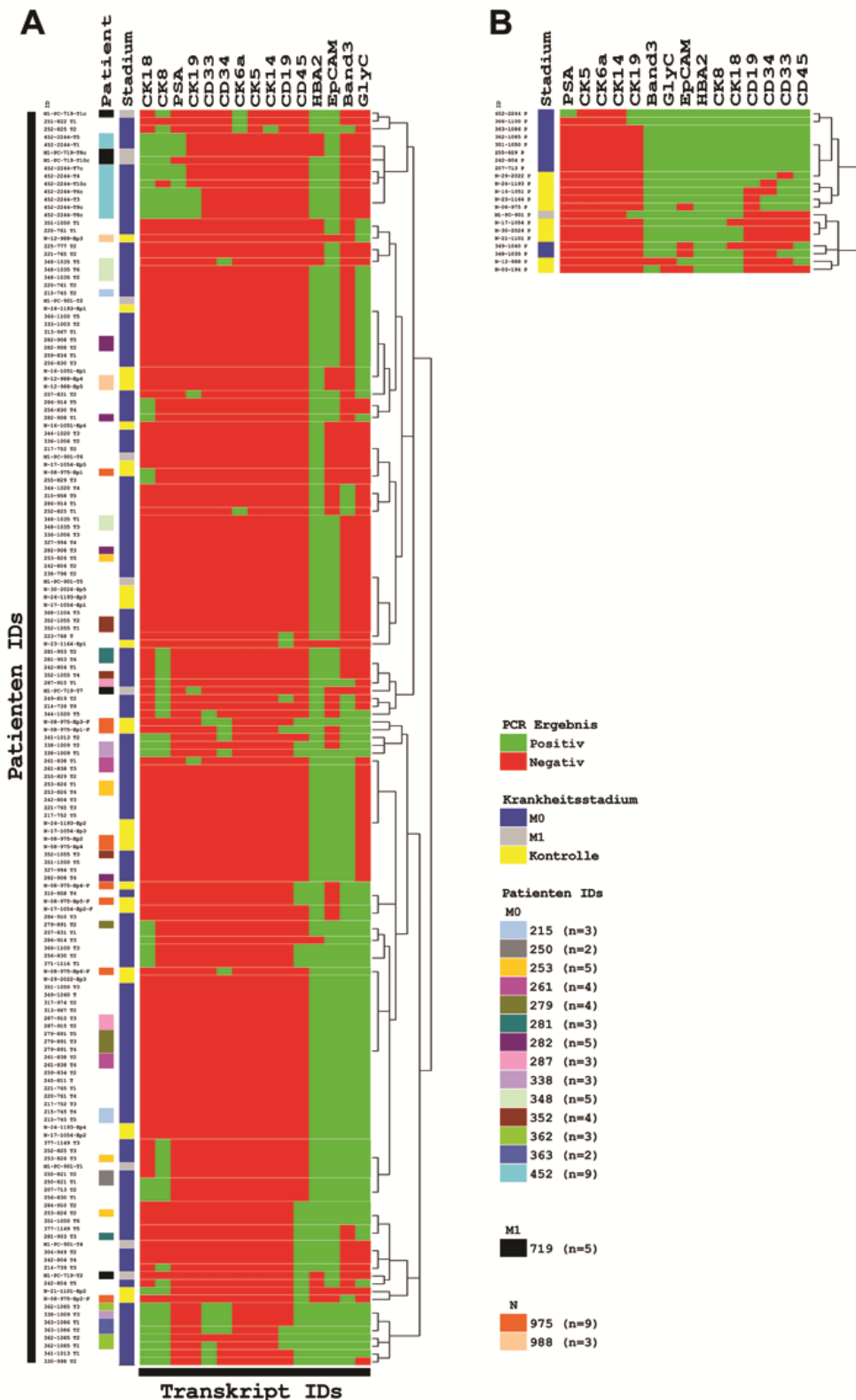


### 3.7 Analyse der genomischen Aberrationen

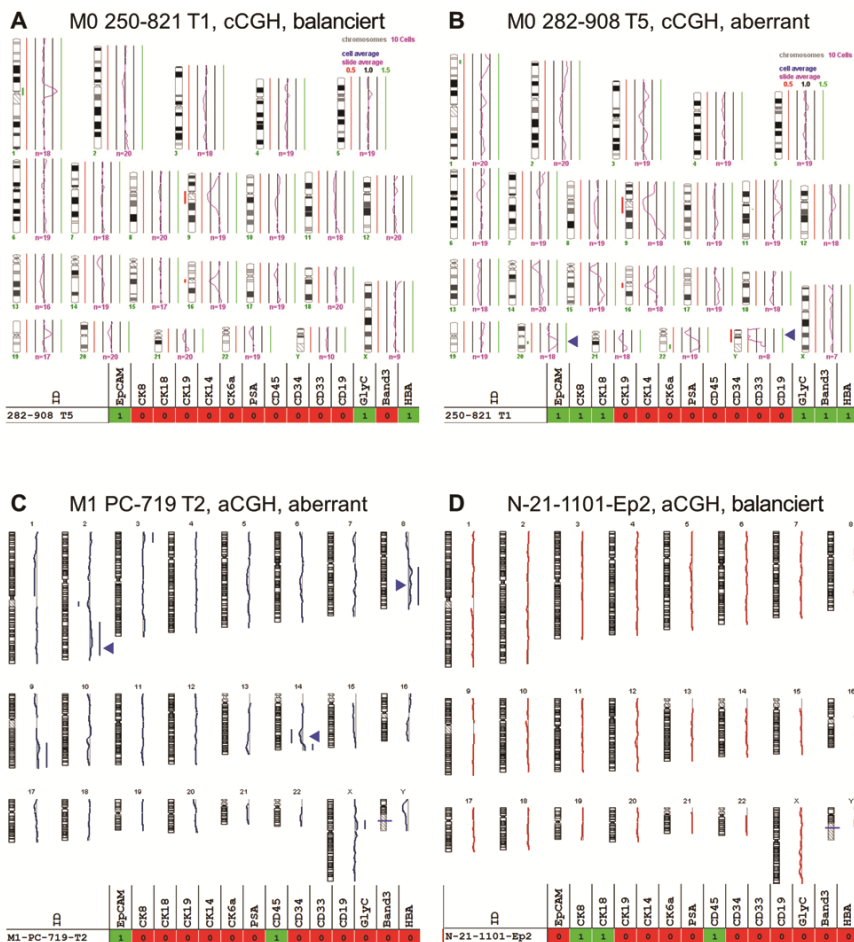
Das Vorhandensein hämatopoetischer Transkripte in EpCAM<sup>+</sup> Einzelzellen aus dem Knochenmark von Prostatakarzinompatienten warf die Frage auf, ob diese trotz der Koexpression von epithelialen Markern tatsächlich Tumorzellen waren. Deshalb wurden ausgewählte Einzelzellen nach dem Vorhandensein genomischer Aberrationen mittels komparativer genomischer Hybridisierung (CGH) überprüft. Dieser Teil der Arbeit wurde von Herrn Dr. Miodrag Guzvic durchgeführt. Das Vorhandensein genomischer Aberrationen beweist den Tumorsprung einer Zelle, wohingegen das Fehlen genomischer Aberrationen aufgrund der limitierten Auflösung der Methode nicht gleichbedeutend mit dessen Ausschluss ist (Schardt et al., 2005).

Um die Verlässlichkeit der kombinierten Genom- und Transkriptomanalyse zu demonstrieren wurde zunächst das CGH Profil einer VCaP Einzelzelle mit dem von unamplifizierter VCaP DNA verglichen. Die Mehrzahl der bekannten genetischen Aberrationen konnten auch im Genom der Einzelzelle identifiziert werden (siehe Supplementary Figure 10A). Im Gegensatz hierzu konnte bei zwei Leukozyten aus dem Blut eines gesunden Spenders (periphere Blutleukozyten, PBLs) ein normaler Karyotyp nachgewiesen werden (Supplementary Figure 10B).

Daraufhin wurden die WGA's von 4 EpCAM<sup>+</sup> Einzelzellen aus der Kontrollgruppe und 34 EpCAM<sup>+</sup> Einzelzellen aus dem Knochenmark von Patienten im M0- und M1-Stadium mittels CGH analysiert. In den vier Zellen der Kontrollgruppe konnten keine genomischen Aberrationen nachgewiesen werden (Supplementary Figure 10C), wohingegen 13 von 34 Zellen (38%) aus dem Knochenmark von Prostatakarzinompatienten genomische Aberrationen aufwiesen. Erstaunlicherweise wurden dabei Einzelzellen gefunden, die sowohl hämatopoetische (z.B. CD45) und speziell erythrozytäre Marker (z.B. *HBA2*) exprimierten als auch gleichzeitig genomische Aberrationen beherbergten. Dies betraf sowohl Zellen aus dem Knochenmark von M0- als auch von M1-Patienten (Abb. 17, Supplementary Figure 10E). Damit konnte nachgewiesen werden, dass DCCs in der Tat hämatopoetische bzw. erythrozytäre Transkripte exprimieren können.

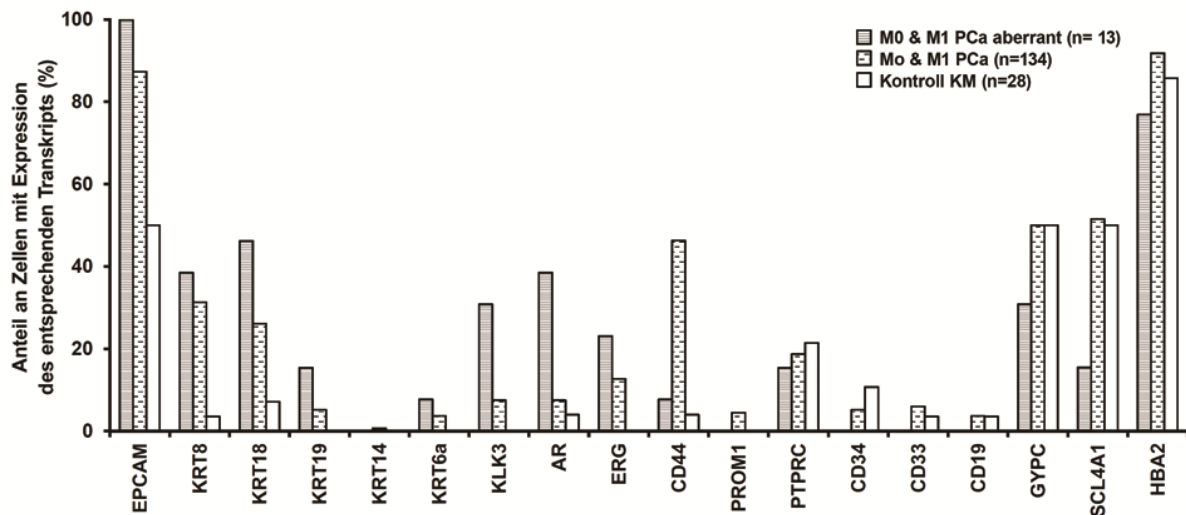


**Abbildung 16.** Hierarchische Clusteranalyse der Transkriptomanalyse aus EpCAM<sup>+</sup> Einzelzellen (A) und Zellpools (B), die aus dem Knochenmark von M0- und M1-PCa Patienten sowie Kontrollpatienten stammen. Die grüne Farbkodierung bedeutet, dass ein entsprechendes Transkript in der jeweiligen Probe mittels PCR detektiert wurde, die rote Farbe signalisiert ein negatives Ergebnis. Die Probenart (M0- bzw. M1-PCa oder N [Kontrolle]) ist ebenfalls farbkodiert. Zellen, die sich miteinander gruppieren und aus dem Knochenmark des gleichen Patienten stammen, wurden ebenfalls farbkodiert. In Abb. 15 sind die Ergebnisse der Transkriptomanalyse noch einmal nach Probenart (Einzelzelle oder Zellpool) und Probenotyp (Tumorpatient und Kontrollpatient) gegliedert.



**Abbildung 17.** Kombinierte Genom- und Transkriptomanalyse von EpCAM<sup>+</sup> Einzelzellen. **A** und **B** zeigen CGH Profile von zwei M0-Prostatakarzinompatienten, **C** einen M1-Patienten und **D** einen Kontrollpatienten. Einige der festgestellten Aberrationen sind durch kleine Pfeile hervorgehoben. Da die Zentromerregion bekanntlich falsch-positive Signale generiert, wurde sie nicht in die Analyse einbezogen. Das Resultat der entsprechenden Genexpressionsanalyse findet sich unter dem jeweiligen CGH Profil. Die weiteren CGH Profile können dem Anhang entnommen werden (Supplementary Fig. S10). cCGH – CGH auf Metaphasen; aCGH – CGH auf DNA Microarrays

Das Expressionsmuster aus den 21 getesteten Genen wurde dann zwischen EpCAM<sup>+</sup> Zellen, die genomische Aberrationen gezeigt hatten, und dem Gesamtkollektiv der EpCAM<sup>+</sup> Zellen aus dem Knochenmark von Prostatakarzinompatienten verglichen. Insgesamt gesehen hat sich das Expressionsmuster in beiden Gruppen kaum unterschieden, wobei unter den Zellen mit genomischen Aberrationen tendenziell mehr Zellen epitheliale Transkripte (*KRT8*, *KRT18*; *KRT19*; *KRT6a*, *EPCAM*) beinhalteten. Die Expression von erythrozytären Markern war in beiden Kohorten in vergleichbarer Weise zu detektieren (z.B. *HBA2* 77% vs. 92% und *CD45* 15% vs. 18%). Der wohl einzige relevante Unterschied betraf die Expression von *KLK3* (PSA) und *AR*. Hier exprimierten die Zellen mit genomischen Aberrationen signifikant häufiger PSA und *AR* als das Gesamtkollektiv ( $p = 0,02$  für PSA und  $p = 0,004$  für *AR*; jeweils Fisher's Exakt Test, Abb. 18).



**Abbildung 18.** Genexpressionsanalyse von disseminierten Tumorzellen aus dem Knochenmark von Prostatakarzinompatienten (PCa). Zellen mit nachgewiesenem malignem Ursprung exprimieren ein ähnliches Profil wie die Gesamtgruppe der EpCAM<sup>+</sup> Einzelzellen aus den Prostatakarzinompatienten. Die jeweiligen Transkripte befinden sich auf der x-Achse. Die Ergebnisse für *KRT5*, *MAGEA2* und *MAGEA4* wurden nicht dargestellt, da diese Transkripte in keiner der Proben detektiert wurden. EpCAM<sup>+</sup> Zellen aus M0- und M1-Patienten (graue Balken), EpCAM<sup>+</sup> Zellen mit genomischen Aberrationen (schwarze Balken) und Kontrollpatienten (weiße Balken). *n* entspricht der Anzahl an untersuchten Einzelzellen.

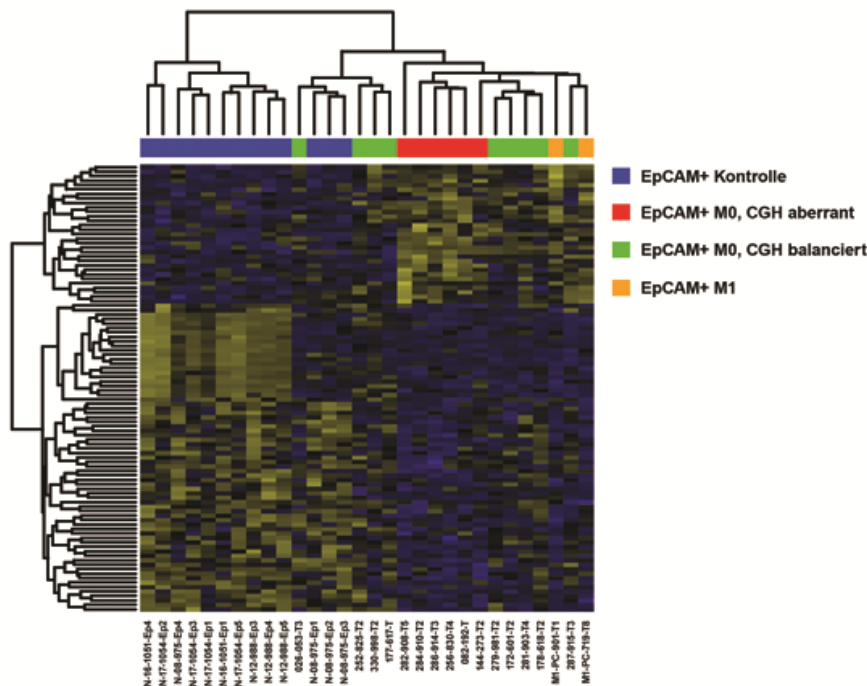
### 3.8. Transkriptomanalyse von EpCAM<sup>+</sup> Einzelzellen mittels Microarrays

Nachdem mittels CGH eindeutig disseminierte Tumorzellen nachgewiesen wurden, sollte nun mittels vergleichender Transkriptomanalyse untersucht werden, inwieweit sich die Transkriptome von EpCAM<sup>+</sup> DCCs aus dem Knochenmark von Prostatakarzinompatienten und EpCAM<sup>+</sup> Einzelzellen aus dem Knochenmark von Kontrollpatienten unterscheiden. Das Ziel ist die Identifikation neuer Gene, die zum einem die Detektierung von DCCs im Knochenmark erleichtern könnten, als auch Rückschlüsse auf biologische Vorgänge in DCCs zulassen, insbesondere aktivierte Signalwege.

Hierzu wurden die WTAs von 6 aberranten EpCAM<sup>+</sup> Einzelzellen aus dem Knochenmark von M0-Patienten, 2 aberranten EpCAM<sup>+</sup> Einzelzellen aus dem Knochenmark von M1-Patienten und 13 EpCAM<sup>+</sup> Einzelzellen aus dem Knochenmark von Kontrollpatienten auf humanen Oligonukleotidarrays hybridisiert. Anhand der Expressionsprofile wurde eine Liste von 100 Genen erstellt, die signifikant unterschiedlich zwischen aberranten Zellen und Kontrollzellen exprimiert wurden. Diese Liste wurde als Klassifikator verwendet. Anhand dessen wurden anschließend 9 EpCAM<sup>+</sup> Einzelzellen aus dem Knochenmark von M0-Patienten mit balanciertem CGH Karyotyp bzw. unbekanntem CGH Profil mit den Expressionsprofilen der aberranten disseminierten Tumorzellen bzw. der Kontrollzellen verglichen und den jeweiligen Gruppen zugeordnet (Abb. 19). Die untersuchten Einzelzellen gruppieren sich hauptsächlich in zwei Gruppen. Eine Gruppe bestand fast ausschließlich aus EpCAM<sup>+</sup> Kontrollzellen, während die zweite Gruppe sowohl die



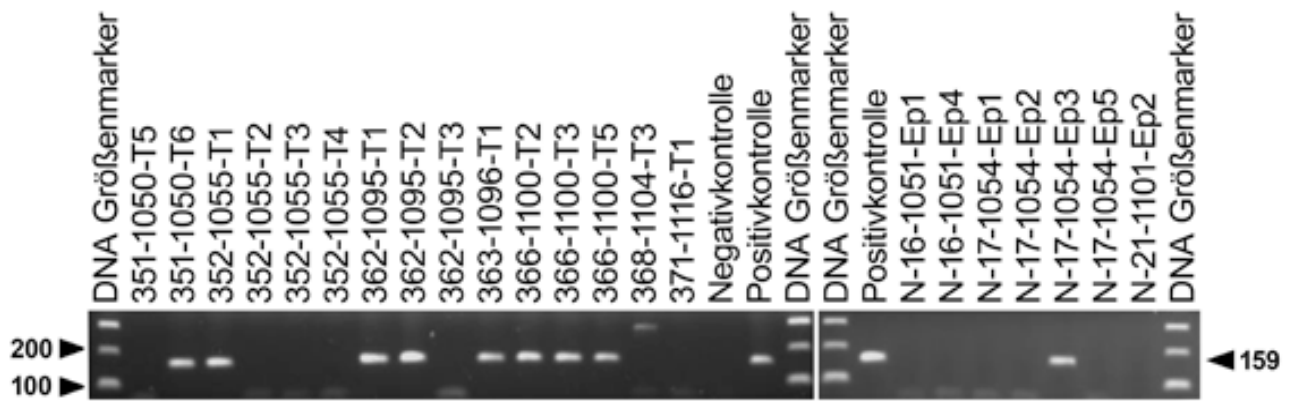
aberranten Einzelzellen der M0- und M1- Patienten als auch den Großteil der EpCAM<sup>+</sup> Zellen aus den M0-Patienten mit balancierten bzw. unbekanntem CGH Profil beinhaltete. Dennoch gruppierten sich auch drei EpCAM<sup>+</sup> Kontrollzellen in eine Subgruppe des zweiten Hauptclusters.



**Abbildung 19.** Einteilung der EpCAM<sup>+</sup> Einzelzellen anhand der Ähnlichkeit ihres Transkriptoms mithilfe eines Klassifikators, der auf signifikant unterschiedlich exprimierten Genen basiert. Vergleichsgruppen waren EpCAM<sup>+</sup> Einzelzellen aus Kontrollpatienten (blaue Gruppe) und aberrante EpCAM<sup>+</sup> Einzelzellen aus dem Knochenmark von M0-Prostatakarzinompatienten (rote Gruppe). EpCAM<sup>+</sup> Einzelzellen aus M1-Patienten und EpCAM<sup>+</sup> Einzelzellen aus M0-Patienten mit balanciertem CGH Profil bzw. unbekanntem Ergebnis klassifizieren hauptsächlich mit den aberranten Einzelzellen. Die dunkle Farbe im „thermischen Mapping“ bezeichnet eine niedrige, die gelbe Farbe eine hohe Genexpression. Die Spalten für die einzelnen Gene sind horizontal angelegt.

Um die Ergebnisse dieser Transkriptomanalyse zu validieren, wurde ein Gen aus der Liste des Klassifikators ausgewählt, dessen Transkripte in EpCAM<sup>+</sup> Zellen aus dem Knochenmark von Prostatakarzinompatienten vorhanden und in EpCAM<sup>+</sup> Zellen aus Kontrollpatienten abwesend sein sollten. Dieses Gen, *DHRS3*, codiert für eine Dehydrogenase/Reduktase als Teil des Retinsäure-Metabolismus (Kirschner et al., 2010). Es wurde mittels analytischer PCR in 52 (42%) EpCAM<sup>+</sup> Zellen aus dem Knochenmark von M0-Patienten detektiert, wohingegen nur eine Zelle aus der Kontrollgruppe (4%) dieses Gen exprimiert ( $p < 0,001$ ; Fisher's Exakt Test). Dennoch war auch mittels *DHRS3* Expression keine klare Trennung von EpCAM<sup>+</sup> Kontrollzellen und EpCAM<sup>+</sup> disseminierten Tumorzellen möglich, da das Vorhandensein von genomischen Aberrationen nicht zu 100% mit dem Vorhandensein von *DHRS3* mRNA korrelierte bzw. nicht alle Zellen der Kontrollpatienten negative PCR-Ergebnisse für *DHRS3* aufwiesen (Abb. 20).





**Abbildung 20.** *DHRS3* Expression in ausgewählten EpCAM<sup>+</sup> Einzelzellen aus PCa Patienten und Kontrollpatienten; Ergebnisse der analytischen PCR. Die Photos eines jeden Gels wurden zurechtgeschnitten, sodass sie die jeweiligen Zielfragmente und die angrenzenden Banden des Größenmarkers enthalten (Pfeilmarkierung). Die Probennummer ist über den Spalten abgebildet. Helligkeit und Kontrast wurden mittels „Auto Kontrast“ Tool des Programms Adobe Photoshop angepasst.

## 4. Diskussion

In dieser Arbeit wurde eine gezielte Transkriptomanalyse an EpCAM<sup>+</sup> Einzelzellen aus dem Knochenmark von nicht-metastasierten (M0-Stadium) Prostatakarzinompatienten durchgeführt und es konnte nachgewiesen werden, dass im Gegensatz zu anfänglichen Erwartungen eine mRNA Signatur aus 21 ausgewählten histogenetischen Marker nicht ausreicht, um DCCs verlässlich zu identifizieren. Mittels kombinierter Genom- und Transkriptomanalyse konnten jedoch gezeigt werden, dass Prostatakarzinom-DCCs das Transkriptom der Knochenmarkszellen nachahmen und hämatopoetische, insbesondere erythrozytäre Transkripte exprimieren. Diese Erkenntnis deutet auf eine phänotypische Plastizität der Tumorzellen hin, die weit über das bekannte Konzept der epithelialen-mesenchymalen Transition hinausgeht (Thierry et al., 2009).

Zu Beginn wurden Zellpools aus dem Knochenmark von Prostatakarzinom- und Kontrollpatienten analysiert, aus denen zuvor CD45-, CD33-, CD11b- und Glycophorin A-positive Zellen entfernt wurden. Die Idee dahinter war, dass ein oder mehrere Marker in der Lage sein sollten die Proben aus Tumorpatienten eindeutig zuzuordnen. Jedoch exprimierten die Zellpools beider Gruppen *EPCAM*, *KRT8*, *KRT18*, hämatopoetische und erythrozytäre Transkripte, was den Ergebnissen publizierter Studien entspricht (de Graaf et al., 1997; Dimmler et al., 2001). *CD19* war in Zellpools aus Tumorpatienten deutlich häufiger exprimiert, ohne dass wir eine passende Erklärung hierfür haben. Da die Analyse der Zellpools keine wesentlichen Erkenntnisse erbrachte, wurde sich auf die Analyse isolierter EpCAM<sup>+</sup> Einzelzellen fokussiert, da diese Population am wahrscheinlichsten aus disseminierten Tumorzellen mit einem charakteristischen Expressionsprofil besteht.

Die Expression des EpCAM Proteins schien in Proben aus Prostatakarzinompatienten stärker zu sein als in Kontrollpatienten. Der Anteil an EpCAM-positiven Proben war jedoch in beiden Gruppen vergleichbar (62% vs. 56%) und eine alleinige Identifikation von DCCs anhand der Qualität der EpCAM-Färbung war nicht sicher möglich (Klein et al., 2002; Choesmel et al., 2004; Watson et al., 2007). Deshalb wurde untersucht, ob wir mit Hilfe eines Marker-Panels bestehend aus epithelialen (*EPCAM*, *KRT8*, *KRT18*, *KRT19*, *KRT14*, *KRT6a*, *KRT5*), prostataassoziierten (*KLK3/PSA*, *AR*, *ERG*, *CD44*, *PROM1/CD133*), tumorassoziierten (*MAGEA2*, *MAGEA4*) und den hämatopoetischen Transkripten *PTPRC* (CD45), *CD33*, *CD34*, *CD19*, *GYPC*, *SCL4A1* (Bande 3 Protein) sowie *HBA2* zwischen EpCAM<sup>+</sup> Zellen aus Prostatakarzinom- und Kontrollpatienten unterschieden werden kann.

Für alle untersuchten Transkripte wurde ein PCR Assay mit Einzelzellsensitivität entwickelt. Trotzdem gelang mit keinem der gewählten Transkripte eine eindeutige DCC Identifikation, da entweder nur sehr wenige DCCs das entsprechende Transkript exprimierten, oder die Kontrollzellen ebenso häufig positiv waren. Mittels hierarchischer Clusteranalyse konnte ebenfalls

kein Expressionsmuster gefunden werden, welches DCCs eindeutig von EpCAM-positiven Kontrollzellen unterscheidet. Trotz dieser unerwarteten Ergebnisse wurden interessante Beobachtungen gemacht.

(I) Auch in den Einzelzellen korreliert das Vorhandensein des EpCAM Transkripts mit der Intensität der Fluoreszenzfärbung. In 88 Prozent der subjektiv heller gefärbten EpCAM<sup>+</sup> Einzelzellen aus dem Knochenmark von Prostatakarzinompatienten konnte ein EpCAM Transkript entdeckt werden, wohingegen dies nur in 50 Prozent der schwächer gefärbten EpCAM<sup>+</sup> Kontrollzellen der Fall war. Dies legt nahe, dass diese Ergebnisse technisch wahrscheinlich sehr verlässlich sind. Da EpCAM ein stark glykosyliertes Protein ist, besitzt es eine lange Halbwertszeit, was die niedrigere Transkriptionsrate in den Kontrollzellen erklären könnte (Munz et al., 2008).

(II) EpCAM<sup>+</sup> Zellen aus dem Knochenmark von Prostatakarzinompatienten koexprimierten signifikant häufiger verschiedene epitheliale Transkripte als EpCAM<sup>+</sup> Kontrollzellen (93% vs. 57%). Dies betraf insbesondere *KRT8* und *KRT18*, welche für Proteine kodieren, die üblicherweise von den in der DCC-Forschung gängigen anti-Zytokeratin Antikörpern, wie dem monoklonalen Antikörper A45-B/B3, erkannt werden (Fehm et al., 2006). Die anderen getesteten Zytokeratintaskripte (*KRT5*, *KRT6a*, *KRT14*) wurden in den EpCAM<sup>+</sup> Zellen der Prostatakarzinompatienten nur sehr selten exprimiert. Diese Beobachtungen unterstützen die Hypothese, dass EpCAM<sup>+</sup> Zellen aus dem Knochenmark der Prostatakarzinompatienten tatsächlich Zellen epithelialen Ursprungs sind und damit DCCs. Die Vermutung, dass disseminierte Tumorzellen Zytokeratine exprimieren, die typischerweise in basalen, stammzellähnlichen Kompartimenten vorkommen, so wie *KRT5*, *KRT6a* und *KRT14* (Yang et al., 1997; van Leenders et al., 2003; Collins et al., 2005; Schmelz et al., 2005), kann jedoch nicht bestätigt werden. Hierfür spricht auch die sehr niedrige Rate an Koexpression von *CD44* und *PROM1*. Deren Vorkommen wurde auf Tumorstammzellen des Prostatakarzinoms beschrieben (Collins et al., 2005; Vander Griend et al., 2008), wobei EpCAM<sup>+</sup>/PROM1<sup>+</sup> Tumorstammzellen bislang nur in Lymphknotenmetastasen und nicht in Knochenmetastasen des Prostatakarzinoms detektiert wurden (Eaton et al., 2010).

(III) Transkripte für die prostataassoziierten Marker PSA, *AR* und *ERG* bzw. tumorassoziierte Transkripte (*MAGEA*) (Boon et al., 1994) wurden in den EpCAM<sup>+</sup> Zellen der Prostatakarzinompatienten nur sehr selten bzw. gar nicht detektiert. Das Fehlen der *MAGEA* Transkripte entspricht am ehesten den niedrigen Detektionsraten in Knochenmarkpools (Kufer et al. 2002; Mecklenburg et al., 2007), das seltene Vorhandensein der PSA Expression lässt sich mit ähnlichen Arbeiten an CK<sup>+</sup> Zellen vergleichen (Riesenberg et al., 1993; Mueller et al., 1998). Trotzdem sei hier zu erwähnen, dass in anderen Studien deutlich höhere Raten an EpCAM<sup>+</sup> DCCs mit PSA Expression gefunden wurden (Ellis et al., 2003; Pfitzenmaier et al., 2007). Aufgrund

unterschiedlicher Methodik ist ein direkter Vergleich der Ergebnisse jedoch schwierig. So wurden die PSA Expression in den genannten Studien nur in gepoolten Proben analysiert. Es kann nicht ausgeschlossen werden, dass mit unserer Einzelzell-basierten Methode PSA exprimierende DCCs „verpasst“ wurden. Da die Expression von PSA in unserem Kollektiv vermehrt mit der Bildung von Zellclustern bzw. Mikrometastasen einherging, bestehen die EpCAM<sup>+</sup> DCCs ohne PSA Expression möglicherweise aus undifferenzierten Zellen. Analog zur niedrigen Expressionsrate von PSA wurden auch für AR nur eine geringe Anzahl an EpCAM<sup>+</sup> Zellen detektiert, die AR exprimierten. Die isolierten EpCAM<sup>+</sup> Zellen aus dem Knochenmark von M1-Patienten wiesen jedoch höhere Raten an PSA und AR Expression auf verglichen mit der Gruppe der M0-Patienten. Dies könnte darauf hindeuten, dass die DCCs den prostatatypischen Phänotyp während der Tumorprogression wiedererlangen. Interessanterweise zeigten die EpCAM<sup>+</sup> Zellen aus dem Knochenmark der M0-Patienten eine signifikant höhere Rate an ERG Expression auf im Vergleich zur Kontrollgruppe. Die Überexpression von ERG in Prostatakarzinomzellen führte zur Entdeckung der prostatakarzinomspezifischen *TMPRSS2/ERG* Fusion (Tomlins et al., 2005). Möglicherweise ist dies ein Indiz, dass einige der EpCAM<sup>+</sup> Einzelzellen aus den M0-Patienten diese tumorspezifische Genfusion beinhalten. Aus technischen Gründen konnte dies jedoch nicht untersucht werden. Da die möglichen Bruchpunkte der Fusion zwischen den einzelnen Patienten stark variieren, bedürfte es einer Sequenzierung der jeweiligen WTAs um eine verlässliche Aussage über das Vorliegen des Fusionstranskriptes zu treffen (Clark et al., 2008; Weier et al., 2013).

Zu guter Letzt war auch die Untersuchung der hämatopoetischen Marker nicht hilfreich DCCs von der autochthonen EpCAM<sup>+</sup> Knochenmarkspopulation zu unterscheiden. Wir konnten keine signifikanten Unterschiede in der Häufigkeit von *PTPRC* (CD45), *CD19*, *CD33* und *CD34* exprimierenden Einzelzellen im M0-Stadium und Kontrollpatienten finden. Da veröffentlichte Studien darauf hinweisen, dass frühe Reifungsstufen der Erythropoese möglicherweise EpCAM exprimieren (Buhring et al., 1996; Lammers et al., 2002), analysierten wir die Expression von Transkripten in den EpCAM<sup>+</sup> Kontrollzellen, die als typisch für die erythrozytäre Reihe gelten. Dies waren *HBA2*, *EPOR* (Israels et al., 2003), *SCL4A1* (Fukuda et al., 1980), und *GYPC* (Villevall et al., 1989), *TFRC* (Buhring et al., 1996), *KIT*, *NFE2* und *KLF1* (Ziegler et al., 1999). Insgesamt 96% dieser Zellen exprimierten mindestens eines dieser Transkripte, wobei *HBA2* das am häufigsten detektierte war (86%). Diese Beobachtung war – auch in Anbetracht des Fehlens anderer hämatopoetischer Marker – ein starkes Indiz dafür, dass die EpCAM<sup>+</sup> Zellen aus den Kontrollpatienten tatsächlich der erythrozytären Reihe entstammen. Als jedoch die *HBA2*, *SCL4A1* und *GYPC* Expression an den EpCAM<sup>+</sup> Zellen aus M0- und M1-Patienten getestet wurde, zeigte sich überraschenderweise, dass 96% der Zellen aus M0-Patienten ebenfalls mindestens eines dieser Gene exprimierten. Auch hier war *HBA2* das am häufigsten vorkommende Gen (93%). Die Rate an Zellen, welche erythrozytäre Transkripte exprimieren, unterschied sich nicht zwischen M0-

und Kontrollpatienten. Auch in den EpCAM<sup>+</sup> Zellen aus den M1-Patienten exprimierten 80% der Zellen *HBA2*.

Trotz der Tatsache, dass sich EpCAM<sup>+</sup> Einzelzellen aus dem Knochenmark von M0- und M1-Prostatakarzinompatienten deutlich von den EpCAM<sup>+</sup> Einzelzellen aus dem Knochenmark der Kontrollpatienten unterscheiden, schränken sich durch die gemachten Beobachtungen die Möglichkeiten erheblich ein, DCCs zweifelsfrei zu identifizieren. Zuerst wurde die Möglichkeit einer Kontamination der WTA durch extrazelluläre mRNA ausgeschlossen. Anschließend haben wir nach einer weiteren Möglichkeit gesucht, die EpCAM<sup>+</sup> Zellen der Prostatakarzinompatienten als DCCs zu identifizieren. Dies wurde durch eine kombinierte Genom- und Transkriptomanalyse in ausgewählten Einzelzellen erreicht. Wir konnten beobachten, dass Zellen mit nachgewiesenem malignen Ursprung häufiger Zytokeratine exprimieren im Vergleich zum Gesamtkollektiv an EpCAM<sup>+</sup> Zellen. Dies deutet darauf hin, dass hämatopoetische EpCAM<sup>+</sup> Zellen den Anteil wirklicher DCCs in der Gruppe der M0-Patienten „verwässert“ haben. Tatsächlich war die Expression von *KLK3* (PSA) und *AR* signifikant erhöht in der Gruppe der genomisch aberranten EpCAM<sup>+</sup> Zellen. Somit bieten diese beiden Marker aktuell die einzige Möglichkeit DCCs des Prostatakarzinoms korrekt zu identifizieren. Es bedarf jedoch weiterer Studien mit größeren Patientenzahlen, da *KLK3* exprimierende Zellen lediglich im Knochenmark von zwei Patienten gefunden wurden und die Zahl der Zellen mit Expression von *AR* nur sehr gering war. Außerdem wiesen viele EpCAM<sup>+</sup> Zellen mit sicherem malignen Ursprung keine Transkripte von *KLK3* und *AR* auf, sondern ein ähnliches Expressionsmuster wie die übrigen EpCAM<sup>+</sup> Zellen der M0-Prostatakarzinompatienten. Die Tatsache, dass viele EpCAM<sup>+</sup> Zellen mit genomischen Aberrationen sowohl hämatopoetische Marker wie z.B. CD45, als auch speziell erythrozytäre Marker exprimieren können, weist auf eine unerwartet hohe phänotypische Plastizität in DCCs hin und betrifft beim Prostatakarzinom sowohl DCCs in den frühen Stadien, als auch disseminierte Tumorzellen aus metastasierten Stadien.

Diese Beobachtungen haben erhebliche Auswirkungen auf zukünftige Studien zum Prozess der frühen Metastasierung. Zunächst einmal scheinen gängige Annahmen über die Expression von DCC-spezifischen Markern wenig hilfreich, DCCs sicher zu identifizieren. Viele Studien vermuten für DCCs eine histogenetische Spezifität (Attard et al., 2009), die – zumindest auf mRNA Ebene – bei Prostatakarzinompatienten nicht bestätigt werden konnte. Somit müssen beispielsweise Versuche, das Tumorzelltranskriptom von CTCs mittels einfacher „Transkriptsubtraktion“ zu rekonstruieren, mit Vorsicht gesehen werden (Yu et al., 2012). Die Beobachtungen zur Expression von hämatopoetischen und speziell erythrozytären Transkripten in DCCs werden durch ältere Studien bestätigt, wie z.B. eine Transkriptomstudie an 15 singulären Prostatakarzinom-CTCs eines metastasierten Patienten, die ebenfalls eine mäßiggradige Expression von *EPCAM*, *PTPRC* (CD45) und *CD34* (Ozkumur et al., 2013) offenbarte. Des Weiteren konnte bereits gezeigt werden,

dass verschiedene epitheliale Brustkrebszelllinien Hämoglobin exprimieren können (Gorr et al., 2011; Capulli et al., 2012) und dass man in der Lage ist, Zelllinien aus Ovarialkarzinomen in erythrozytäre Zellen mit Hämoglobinexpression zu transdifferenzieren (Zhang et al., 2013). Die nun beobachtete Expression von hämatopoetischen Transkripten in DCCs aus dem Knochenmark von Prostatakarzinompatienten könnte auf einen Anpassungsmechanismus von epithelialen Zellen an die Umgebung des Knochenmarks hindeuten. Es wurde bereits gezeigt, dass metastasierte Prostatakarzinomzellen mit hämatopoetischen Stammzellen (HSCs) um die Stammzellnische im Knochenmark konkurrieren (Shiozawa et al., 2011). Sollten sich DCCs wirklich in der hämatopoetischen Stammzellnische niederlassen, ist es durchaus möglich, dass mittels zellvermittelten Kontakt die Expression von hämatopoetischen Transkripten in DCCs initiiert wird. Dies könnte durch die ansässigen Knochenmarkszellen, z.B. über direkte Zell-Zell Interaktionen oder sezernierte Faktoren geschehen. Alternativ oder zusätzlich könnte das Transkriptom der DCCs durch Aufnahme von mRNA-haltigen Exosomen verändert werden, die von den ansässigen Knochenmarkszellen sezerniert werden (Ono et al., 2014). Auch könnten die gemachten Beobachtungen Folge von hypoxischem oder oxidativem Stress sein, dem DCCs im Knochenmark unterworfen sind (Liu et al., 2011; Emara et al., 2014). Weitere Studien sind notwendig, um die zugrundeliegenden Ursachen unserer Beobachtungen zu identifizieren. Die im Vergleich zu bekannten Prostatakarzinomzellen (Attard et al., 2009) veränderten Transkriptome der DCCs führen wahrscheinlich zu neuartigen Phänotypen mit Ausbildung unterschiedlicher Zellfunktionen. Dies muss bei der Entwicklung von gezielten Tumorthherapie beachtet werden.

Eine ausführliche vergleichende Genexpressionsanalyse von EpCAM<sup>+</sup> DCCs aus Prostatakarzinompatienten und EpCAM<sup>+</sup> Zellen aus Kontrollpatienten wird helfen, neue, spezifische Marker zu identifizieren und die Detektierung von DCCs im Knochenmark erleichtern. Zudem können auch genauere Einblicke in biologische Vorgänge in DCCs gewonnen werden, insbesondere aktivierte Signalwege. Mittels der von uns durchgeführten kombinierten Genom- und Transkriptomanalyse konnten an einem kleinen Testkollektiv bereits einige signifikant unterschiedlich exprimierte Gene in EpCAM<sup>+</sup> DCCs identifiziert werden. Es fiel auf, dass in den Tumorzellen insbesondere Gene hochreguliert waren, die im Zusammenhang mit DNA-Reparaturmechanismen, Chromosomenstrukturierung, transkriptioneller Aktivität und Aktivierung des mitochondrialen Metabolismus stehen. Dies könnte eine Reaktion der Tumorzellen auf die ungewohnte Umgebung des Knochenmarks darstellen. Eine Validierung dieser Ergebnisse gelang für eines der identifizierten Gene, *DHRS3*. Dennoch war auch hier keine klare Trennung von EpCAM<sup>+</sup> Kontrollzellen und EpCAM<sup>+</sup> DCCs möglich, da das Vorhandensein von genomischen Aberrationen nicht zu 100% mit dem Vorhandensein von *DHRS3* mRNA korrelierte. Dies mag dem relativ kleinem Testkollektiv geschuldet sein, oder an technischen Limitationen der gewählten Untersuchungsmethoden liegen (Auflösungsgrenzen der komparativen genomischen Hybridisierung und der analytischen PCR).

Bislang ist nicht sehr viel über *DHRS3* bekannt. Es wird wohl durch das Tumorsuppressorgen p53 aktiviert (Kirschner et al., 2010; Deisenroth et al., 2011) und ist mit seinem Proteinprodukt als Reduktase von trans-Retinal zu trans-Retinol am Retinsäuremetabolismus beteiligt. Des Weiteren wurde seine Expression im Zusammenhang mit papillären Schilddrüsenkarzinomen (Oler et al., 2008) als prädiktiver Marker für die Tumorprogression beschrieben. Aufgrund der geringen Größe des untersuchten Kollektivs sind die genannten Ergebnisse sicherlich nur mit Vorsicht zu betrachten. Um verlässliche Marker zu identifizieren muss diese – prinzipiell funktionierende – Methode an einem größeren Kollektiv getestet und anschließend ausreichend validiert werden.

Zusammenfassend lässt sich sagen, dass der Nachweis genomischer Aberrationen weiterhin den alleinigen Goldstandard zur Identifikation von DCCs darstellt. Die kombinierte Genom- und Transkriptomanalyse von DCCs bietet jedoch einen vielversprechenden Ansatz, disseminierte Tumorzellen sicher zu identifizieren und ihren Phänotyp zu beschreiben. Dies wird helfen, die frühe Metastasierung besser zu verstehen und die Entwicklung von Tumortherapien voranzutreiben, die gezielt die Entstehung von Metastasen verhindern können.

## 5. Zusammenfassung

Die ossäre Metastasierung ist beim Prostatakarzinom (PCa) die häufigste Art der Metastasenbildung und Patienten mit ossärer Metastasierung werden für unheilbar erachtet. Die Entwicklung gezielter Therapie gegen Prostatakarzinomzellen, die bereits vor Prostataktomie ins Knochenmark disseminiert sind – d.h. vor der klinisch fassbaren Metastasierung – könnte einen Ansatz darstellen, den letztendlich letal verlaufenden Prozess der Metastasierung zu verhindern. Deshalb analysierten wir das Transkriptom von disseminierten Tumorzellen (DCCs), die aus dem Knochenmark von Prostatakarzinompatienten ohne manifeste Metastasen (UICC Stadium M0) isoliert wurden. Wir untersuchten 105 Knochenmarksproben von M0-PCa Patienten und 18 Knochenmarksproben von männlichen Kontrollpatienten nach dem Vorhandensein von EpCAM<sup>+</sup> Einzelzellen. Insgesamt konnten wir 270 Zellen aus beiden Kollektiven isolieren und deren Transkriptome für weitere Untersuchungen amplifizieren. Mittels gezielter Expressionsanalyse sollten DCCs eindeutig identifiziert werden, um sie im Folgenden einer kombinierten Genom- und Transkriptomanalyse zuzuführen. Alle Zellen wurden auf die Expression von *EPCAM*, *KRT8*, *KRT18*, *KRT19*, *KRT14*, *KRT6a*, *KRT5*, *KLK3* (PSA), *AR*, *ERG*, *CD44*, *PROM1* (CD133), *MAGEA2*, *MAGEA4*, *PTPRC* (CD45), *CD33*, *CD34*, *CD19*, *GYPC*, *SCL4A1* (Bande 3 Protein), und *HBA2* untersucht. Mittels dieser Transkripte war es jedoch unmöglich DCCs verlässlich zu identifizieren. Durch kombinierte Genom- und Transkriptomanalyse konnten wir nachweisen, dass EpCAM<sup>+</sup> Zellen aus Kontrollpatienten Transkripte exprimierten, die eigentlich epithelialen Zellen zugeordnet werden, wohingegen DCCs auch hämatopoetische Transkripte exprimieren können. Diese Ergebnisse weisen auf eine unerwartet hohe Plastizität epithelialer Tumorzellen im Knochenmark hin und hinterfragen die heutzutage üblichen Kriterien bei der Identifikation von DCCs.



## 6. Literaturverzeichnis

- Allard WJ, Matera J, Miller MC, Repollet M, Connelly MC, Rao C, et al. (2004). Tumor cells circulate in the peripheral blood of all major carcinomas but not in healthy subjects or patients with nonmalignant diseases. *Clin Cancer Res.* 10, 6897-904.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. (1990). Basic local alignment search tool. *J Mol Biol.* 215, 403-10.
- Attard G, Swennenhuis JF, Olmos D, Reid AH, Vickers E, A'Hern R, et al. (2009). Characterization of ERG, AR and PTEN gene status in circulating tumor cells from patients with castration-resistant prostate cancer. *Cancer Res.* 69, 2912-8.
- Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. (2010). GenBank. *Nucleic Acids Res.* 38, D46-D51.
- Boon T, Cerottini JC, Van den Eynde B, van der Bruggen P, Van Pel A. (1994). Tumor antigens recognized by T lymphocytes. *Annu Rev Immunol.* 12, 337-65.
- Brabletz T, Lyden D, Steeg PS, Werb Z. (2013). Roadblocks to translational advances on metastasis research. *Nat Med.* 19, 1104-9.
- Braun S, Vogl FD, Naume B, Janni W, Osborne MP, Coombes RC, et al. (2005). A pooled analysis of bone marrow micrometastasis in breast cancer. *N Engl J Med.* 353, 793-802.
- Buhring HJ, Muller T, Herbst R, Cole S, Rappold I, Schuller W, et al. (1996). The adhesion molecule E-cadherin and a surface antigen recognized by the antibody 9C4 are selectively expressed on erythroid cells of defined maturational stages. *Leukemia.* 10, 106-16.
- Capulli M, Angelucci A, Driouch K, Garcia T, Clement-Lacroix P, Martella F, et al. (2012). Increased expression of a set of genes enriched in oxygen binding function discloses a predisposition of breast cancer bone metastases to generate metastasis spread in multiple organs. *J Bone Miner Res.* 27, 2387-98.
- Choesmel V, Anract P, Hoifodt H, Thiery JP, Blin N. (2004). A relevant immunomagnetic assay to detect and characterize epithelial cell adhesion molecule-positive cells in bone marrow from patients with breast carcinoma: immunomagnetic purification of micrometastases. *Cancer.* 101, 693-703.

- Clark J, Attard G, Jhavar S, Flohr P, Reid A, De-Bono J, et al. (2008). Complex patterns of ETS gene alteration arise during cancer development in the human prostate. *Oncogene*. 27, 1993-2003.
- Collins AT, Berry PA, Hyde C, Stower MJ, Maitland NJ. (2005). Prospective identification of tumorigenic prostate cancer stem cells. *Cancer Res*. 65, 10946-51.
- Czyz ZT, Hoffmann M, Schlimok G, Polzer B, Klein CA. (2014). Reliable Single Cell Array CGH for Clinical Samples. *PLoS One*. 9, e85907.
- Deisenroth C, Itahana Y, Tollini L, Jin A, Zhang Y. (2011). p53-Inducible DHRS3 is an endoplasmic reticulum protein associated with lipid droplet accumulation. *J Biol Chem*. 286, 28343-56.
- de Graaf H, Maelandsmo GM, Ruud P, Forus A, Oyjord T, Fodstad O, et al. (1997). Ectopic expression of target genes may represent an inherent limitation of RT-PCR assays used for micrometastasis detection: studies on the epithelial glycoprotein gene EGP-2. *Int J Cancer*. 72, 191-6.
- Dimmler A, Gerhards R, Betz C, Gunther K, Reingruber B, Horbach T, et al. (2001). Transcription of cytokeratins 8, 18, and 19 in bone marrow and limited expression of cytokeratins 7 and 20 by carcinoma cells: inherent limitations for RT-PCR in the detection of isolated tumor cells. *Lab Invest*. 81, 1351-61.
- Eaton CL, Colombel M, van der Pluijm G, Cecchini M, Wetterwald A, Lippitt J, et al. (2010). Evaluation of the frequency of putative prostate cancer stem cells in primary and metastatic prostate cancer. *Prostate*. 70, 875-82.
- Eisen MB, Spellman PT, Brown PO, Botstein D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA*. 95, 14863-8.
- Ellis WJ, Pfitzenmaier J, Colli J, Arfman E, Lange PH, Vessella RL. (2003). Detection and isolation of prostate cancer cells from peripheral blood and bone marrow. *Urology*. 61, 277-81.
- Emara M, Turner AR, Allalunis-Turner J. (2014). Hypoxia differentially upregulates the expression of embryonic, fetal and adult hemoglobin in human glioblastoma cells. *Int J Oncol*. 44, 950-8.
- Fehm T, Braun S, Muller V, Janni W, Gebauer G, Marth C, et al. (2006). A concept for the standardized detection of disseminated tumor cells in bone marrow from patients with primary breast cancer and its clinical implementation. *Cancer*. 107, 885-92.

- Fukuda M, Fukuda MN, Papayannopoulou T, Hakomori S. (1980). Membrane differentiation in human erythroid cells: unique profiles of cell surface glycoproteins expressed in erythroblasts in vitro from three ontogenic stages. *Proc Natl Acad Sci USA*. 77, 3474-8.
- Gorr TA, Wichmann D, Pilarsky C, Theurillat JP, Fabrizius A, Laufs T, et al. (2011). Old proteins - new locations: myoglobin, haemoglobin, neuroglobin and cytoglobin in solid tumours and cancer cells. *Acta Physiol (Oxf)*. 202, 563-81.
- Hartmann CH, Klein CA. (2006). Gene expression profiling of single cells on large-scale oligonucleotide arrays. *Nucleic Acids Res*. 34, e143.
- Heidenreich A, Bastian PJ, Bellmunt J, Bolla M, Joniau S, Mason MD, et al. (2012) Guidelines on Prostate Cancer: European Association of Urology 2012.
- Israels LG, Israels ID. (2003). Erythropoiesis: an overview. In: Molineux G, Foote MA, Elliott SG, editors. *Erythropoietins and Erythropoiesis*. Basel: Birkhäuser Verlag; p. 3-14.
- Kirschner RD, Rother K, Müller GA, Engeland K. (2010) The retinal dehydrogenase/reductase retSDR1/DHRS3 gene is activated by p53 and p63 but not by mutants derived from tumors or EEC/ADULT malformation syndromes. *Cell Cycle*. 9, 2177-88.
- Klein CA, Schmidt-Kittler O, Schardt JA, Pantel K, Speicher MR, Riethmuller G. (1999). Comparative genomic hybridization, loss of heterozygosity, and DNA sequence analysis of single cells. *Proc Natl Acad Sci USA*. 96, 4494-9.
- Klein CA, Blankenstein TJ, Schmidt-Kittler O, Petronio M, Polzer B, Stoecklein NH, et al. (2002). Genetic heterogeneity of single disseminated tumour cells in minimal residual cancer. *Lancet*. 360, 683-9.
- Klein CA, Seidl S, Petat-Dutter K, Offner S, Geigl JB, Schmidt-Kittler O, et al. (2002). Combined transcriptome and genome analysis of single micrometastatic cells. *Nat Biotechnol*. 20, 387-92.
- Klein CA. (2003). The systemic progression of human cancer: a focus on the individual disseminated cancer cell--the unit of selection. *Adv Cancer Res* 89, 35-67.
- Kufer P, Zippelius A, Lutterbuse R, Mecklenburg I, Enzmann T, Montag A, et al. (2002). Heterogeneous expression of MAGE-A genes in occult disseminated tumor cells: a novel multimarker reverse transcription-polymerase chain reaction for diagnosis of micrometastatic disease. *Cancer Res*. 62, 251-61.

- Lammers R, Giesert C, Grunebach F, Marxer A, Vogel W, Buhring HJ. (2002). Monoclonal antibody 9C4 recognizes epithelial cellular adhesion molecule, a cell surface antigen expressed in early steps of erythropoiesis. *Exp Hematol.* 30, 537-45.
- Liu AY, True LD. (2002) Characterization of prostate cell types by CD cell surface molecules. *Am J Pathol.* 160, 37-43.
- Liu W, Baker SS, Baker RD, Nowak NJ, Zhu L. (2011). Upregulation of hemoglobin expression by oxidative stress in hepatocytes and its implication in nonalcoholic steatohepatitis. *PLoS One.* 6, e24363.
- Maglott D, Ostell J, Pruitt KD, Tatusova T. (2011). Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res.* 39, D52-7.
- Mecklenburg I, Weckermann D, Zippelius A, Schoberth A, Petersen S, Prang N, et al. (2007). A multimarker real-time RT-PCR for MAGE-A gene expression allows sensitive detection and quantification of the minimal systemic tumor load in patients with localized cancer. *J Immunol Methods.* 323, 180-93.
- Morgan TM, Lange PH, Porter MP, Lin DW, Ellis WJ, Gallaher IS, et al. (2009). Disseminated tumor cells in prostate cancer patients after radical prostatectomy and without evidence of disease predicts biochemical recurrence. *Clin Cancer Res.* 15, 677-83.
- Mueller P, Carroll P, Bowers E, Moore D, 2nd, Cher M, Presti J, et al. (1998). Low frequency epithelial cells in bone marrow aspirates from prostate carcinoma patients are cytogenetically aberrant. *Cancer.* 83, 538-46.
- Munz M, Fellingner K, Hofmann T, Schmitt B, Gires O. (2008). Glycosylation is crucial for stability of tumour and cancer stem cell antigen EpCAM. *Front Biosci.* 13, 5195-201.
- Oler G, Camacho CP, Hojaj FC, Michaluart P Jr, Riggins GJ, Cerutti JM. (2008). Gene expression profiling of papillary thyroid carcinoma identifies transcripts correlated with BRAF mutational status and lymph node metastasis. *Clin Cancer Res.* 14, 4735-42.
- Ono M, Kosaka N, Tominaga N, Yoshioka Y, Takeshita F, Takahashi RU, et al. (2014). Exosomes from bone marrow mesenchymal stem cells contain a microRNA that promotes dormancy in metastatic breast cancer cells. *Sci Signal.* 7, ra63.
- Ozkumur E, Shah AM, Ciciliano JC, Emmink BL, Miyamoto DT, Brachtel E, et al. (2013). Inertial focusing for tumor antigen-dependent and -independent sorting of rare circulating tumor cells. *Sci Transl Med.* 5, 179ra47.

- Pfitzenmaier J, Ellis WJ, Hawley S, Arfman EW, Klein JR, Lange PH, et al. (2007). The detection and isolation of viable prostate-specific antigen positive epithelial cells by enrichment: a comparison to standard prostate-specific antigen reverse transcriptase polymerase chain reaction and its clinical relevance in prostate cancer. *Urol Oncol.* 25, 214-20.
- Riesenberg R, Oberneder R, Kriegmair M, Epp M, Bitzer U, Hofstetter A, et al. (1993). Immunocytochemical double staining of cytokeratin and prostate specific antigen in individual prostatic tumour cells. *Histochemistry.* 99, 61-6.
- Rozen S, Skaletsky HJ. (2000). Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S, editors. *Bioinformatics Methods and Protocols.* Totowa, USA: Humana Press; p. 365-86.
- Schardt JA, Meyer M, Hartmann CH, Schubert F, Schmidt-Kittler O, Fuhrmann C, et al. (2005). Genomic analysis of single cytokeratin-positive cells from bone marrow reveals early mutational events in breast cancer. *Cancer Cell.* 8, 227-39.
- Schmelz M, Moll R, Hesse U, Prasad AR, Gandolfi JA, Hasan SR, et al. (2005). Identification of a stem cell candidate in the normal human prostate gland. *Eur J Cell Biol.* 84,341-54.
- Shiozawa Y, Pedersen EA, Havens AM, Jung Y, Mishra A, Joseph J, et al. (2011). Human prostate cancer metastases target the hematopoietic stem cell niche to establish footholds in mouse bone marrow. *J Clin Invest.* 121, 1298-312.
- Stoecklein NH, Hosch SB, Bezler M, Stern F, Hartmann CH, Vay C, et al. (2008). Direct genetic analysis of single disseminated cancer cells for prediction of outcome and therapy selection in esophageal cancer. *Cancer Cell.* 13, 441-53.
- Thiery JP, Acloque H, Huang RY, Nieto MA. (2009). Epithelial-mesenchymal transitions in development and disease. *Cell.* 139, 871-90.
- Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, et al. (2005). Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science.* 310, 644-8.
- Vander Griend DJ, Karthaus WL, Dalrymple S, Meeker A, DeMarzo AM, Isaacs JT. (2008). The role of CD133 in normal human prostate stem cells and malignant cancer-initiating cells. *Cancer Res.* 68, 9703-11.
- van Leenders GJ, Schalken JA. (2003). Epithelial cell differentiation in the human prostate epithelium: implications for the pathogenesis and therapy of prostate cancer. *Crit Rev Oncol Hematol.* 46 Suppl, S3-10.

- VerMilyea MD, Maneck M, Yoshida N, Blochberger I, Suzuki E, Suzuki T, et al. (2011). Transcriptome asymmetry within mouse zygotes but not between early embryonic sister blastomeres. *EMBO J.* 30, 1841-51.
- Villeval JL, Le Van Kim C, Bettaieb A, Debili N, Colin Y, el Maliki B, et al. (1989). Early expression of glycophorin C during normal and leukemic human erythroid differentiation. *Cancer Res.* 49, 2626-32.
- Watson MA, Ylagan LR, Trinkaus KM, Gillanders WE, Naughton MJ, Weilbaecher KN, et al. (2007). Isolation and molecular profiling of bone marrow micrometastases identifies TWIST1 as a marker of early tumor relapse in breast cancer patients. *Clin Cancer Res.* 13, 5001-9.
- Weckermann D, Polzer B, Ragg T, Blana A, Schlimok G, Arnholdt H, et al. (2009). Perioperative activation of disseminated tumor cells in bone marrow of patients with prostate cancer. *J Clin Oncol.* 27, 1549-56.
- Weier C, Haffner MC, Mosbrugger T, Esopi DM, Hicks J, Zheng Q, et al. (2013). Nucleotide resolution analysis of TMPRSS2 and ERG rearrangements in prostate cancer. *J Pathol.* 230, 174-83.
- Yang Y, Hao J, Liu X, Dalkin B, Nagle RB. (1997). Differential expression of cytokeratin mRNA and protein in normal prostate, prostatic intraepithelial neoplasia, and invasive carcinoma. *Am J Pathol.* 150, 693-704.
- Yu M, Ting DT, Stott SL, Wittner BS, Oszlak F, Paul S, et al. (2012). RNA sequencing of pancreatic circulating tumour cells implicates WNT signalling in metastasis. *Nature.* 487, 510-3.
- Zhang S, Mercado-Urbe I, Liu J. (2013). Generation of erythroid cells from fibroblasts and cancer cells in vitro and in vivo. *Cancer Lett.* 333, 205-12.
- Ziegler BL, Müller R, Valtieri M, Lamping CP, Thomas CA, Gabbianelli M, et al. (1999). Unicellular-unilineage erythropoietic cultures: molecular analysis of regulatory gene expression at sibling cell level. *Blood.* 93, 3355-68.

## 7. Wissenschaftliche Originalarbeit und Supplementary Material

Beiträge der einzelnen Autoren zu “*Combined genome and transcriptome analysis of single disseminated cancer cells from bone marrow of prostate cancer patients reveals unexpected transcriptomes*”

**Konzeption und Studiendesign:** M. Gužvić, B. Braun, B. Polzer, C.A. Klein

**Methodenentwicklung:** M. Gužvić, B. Braun

**Datengewinnung:**

*Patientenproben:* R. Ganzer, M. Burger, M. Nerlich, S. Winkler

*Probenaufarbeitung:* M. Gužvić, B. Braun

*Transkriptomanalyse:* M. Gužvić, B. Braun,

*Durchflusszytometrie:* M. Werner-Klein, M. Gužvić

*Genomanalyse:* M. Gužvić, Z. T. Czyż

**Datenauswertung und -interpretation (z.B. statistische Analysen, bioinformatische Analysen):** M. Gužvić, B. Braun, M. Werner-Klein, Z. T. Czyż, B. Polzer

**Verfassen und Korrekturlesen des Manuskripts:** M. Gužvić, B. Braun, M. Werner-Klein, Z. T. Czyż, B. Polzer, C.A. Klein

**Studienleiter:** C.A. Klein

## Combined Genome and Transcriptome Analysis of Single Disseminated Cancer Cells from Bone Marrow of Prostate Cancer Patients Reveals Unexpected Transcriptomes

Miodrag Gužvić<sup>1</sup>, Bernhard Braun<sup>1,2</sup>, Roman Ganzer<sup>3</sup>, Maximilian Burger<sup>3</sup>, Michael Nerlich<sup>4</sup>, Sebastian Winkler<sup>5</sup>, Melanie Werner-Klein<sup>6</sup>, Zbigniew T. Czyż<sup>7</sup>, Bernhard Polzer<sup>7</sup>, and Christoph A. Klein<sup>1,7</sup>

### Abstract

Bone is the most frequent site of metastasis in prostate cancer and patients with bone metastases are deemed incurable. Targeting prostate cancer cells that disseminated to the bone marrow before surgery and before metastatic outgrowth may therefore prevent lethal metastasis. This prompted us to directly analyze the transcriptome of disseminated cancer cells (DCC) isolated from patients with nonmetastatic (UICC stage M0) prostate cancer. We screened 105 bone marrow samples of patients with M0-stage prostate cancer and 18 bone marrow samples of patients without malignancy for the presence of EpCAM<sup>+</sup> single cells. In total, we isolated 270 cells from both groups by micromanipulation and globally amplified their mRNA. We used targeted transcriptional profiling to unambiguously identify DCCs for subsequent in-depth analysis. Transcriptomes of all cells were examined for the expression of *EPCAM*, *KRT8*, *KRT18*, *KRT19*, *KRT14*, *KRT6a*, *KRT5*, *KLK3* (PSA), *MAGEA2*, *MAGEA4*, *PTPRC* (CD45), *CD33*, *CD34*, *CD19*, *GYPC*, *SCL4A1* (band 3), and *HBA2*. Using these transcripts, we found it impossible to reliably identify true DCCs. We then applied combined genome and transcriptome analysis of single cells and found that EpCAM<sup>+</sup> cells from controls expressed transcripts thought to be epithelial-specific, whereas true DCCs may express hematopoietic transcripts. These results point to an unexpected transcriptome plasticity of epithelial cancer cells in bone marrow and question common transcriptional criteria to identify DCCs. *Cancer Res*; 74(24): 7383–94. ©2014 AACR.

### Introduction

Despite substantial investment into cancer research, metastasis—the cause of more than 90% of cancer-related deaths (1)—remains poorly understood. Delineating early steps of metastasis before detection of manifest lesions by clinical imaging techniques may provide chances to prevent lethal outcome. After a short period of circulation in the blood (when they are termed circulating tumor cells, CTC), disseminating cancer cells

lodge in distant organs and are termed disseminated cancer cells (DCC). Obviously, cancer cells remaining in the body of M0-stage patients after surgical removal of the primary tumor comprise founders of later arising lethal metastases and therefore DCCs are the target population of systemic adjuvant therapies. In support of this notion are numerous reports on the prognostic impact of DCCs, which are commonly detected by the histogenetic markers, EpCAM or cytokeratins (CK), in lymph nodes and bone marrow, respectively. In contrast to CTCs, which are mostly detected and studied in patients with metastases (2), DCCs are clearly relevant in early disease stages (3–5).

We previously showed that the presence of CK-positive (CK<sup>+</sup>) DCCs in the bone marrow of M0-stage prostate cancer patients is associated with shorter metastasis-free survival (5). Genetic studies of isolated CK<sup>+</sup> DCCs demonstrated marked heterogeneity of M0-stage DCC genomes within and between patients (5, 6). However, little is known about phenotypic characteristics of DCCs. Survival data suggest the existence of DCCs with different metastasis-initiating potential, as not all patients with DCCs develop metastasis. Furthermore, the fact that some patients without CK<sup>+</sup> cells in bone marrow develop metastasis points to the existence of subpopulations of DCCs undetected by current markers. This is particularly true for prostate cancer where the increase of prostate-specific antigen (PSA) after radical prostatectomy (biochemical relapse) is not associated with increasing numbers of CK<sup>+</sup> cells in bone

<sup>1</sup>Chair of Experimental Medicine and Therapy Research, University of Regensburg, Regensburg, Germany. <sup>2</sup>Department of Oncology, Hospital Barmherzige Brüder, Regensburg, Germany. <sup>3</sup>Department of Urology, Caritas-Hospital St. Josef, University of Regensburg, Regensburg, Germany. <sup>4</sup>Department of Trauma Surgery, University Hospital Regensburg, Regensburg, Germany. <sup>5</sup>Department of Orthopaedic Surgery, University of Regensburg, Bad Abbach, Germany. <sup>6</sup>Department of Immunology, University of Regensburg, Regensburg, Germany. <sup>7</sup>Project Group Personalized Tumour Therapy, Fraunhofer Institute of Experimental Medicine and Toxicology, Regensburg, Germany.

**Note:** Supplementary data for this article are available at Cancer Research Online (<http://cancerres.aacrjournals.org/>).

M. Gužvić and B. Braun contributed equally to this article.

**Corresponding Author:** Christoph A. Klein, University of Regensburg, Franz-Josef-Strauss-Allee 11, Regensburg 93053, Germany. Phone: 499-4194-46720; Fax: 499-4194-46719; E-mail: christoph.klein@ukr.de

doi: 10.1158/0008-5472.CAN-14-0934

©2014 American Association for Cancer Research.



marrow (5). Therefore, targeting metastasis founder cells requires detailed molecular characterization of DCCs. One hindrance to such characterization is the low frequency of M0-stage DCCs, with a median number of 1 to 2 DCCs per  $10^6$  bone marrow cells of positive patients.

Previously, we established an approach to comprehensively characterize the transcriptome of a single cell using microarrays (7, 10). As cytokeratins are intracellular proteins, comprehensive analysis of a single DCC transcriptome requires use of a surface marker such as EpCAM for their detection. Furthermore, detection of EpCAM-positive (EpCAM<sup>+</sup>) cells in bone marrow has also been associated with prostate cancer progression (8). However, unlike CK<sup>+</sup> cells (9), EpCAM<sup>+</sup> cells are often found in the bone marrow of healthy individuals (10, 11), thereby confounding molecular studies of DCCs. Hence, there is an urgent need for an assay to reliably select true DCCs for transcriptome analysis using microarrays. Here, our aim was to detect and isolate EpCAM<sup>+</sup> cells from the bone marrow of M0-stage prostate cancer patients and healthy controls for detailed characterization of their gene expression. The objective of this study was to examine the expression of a panel of potentially informative genes that might discriminate individual EpCAM<sup>+</sup> cells found in bone marrow of prostate cancer from EpCAM<sup>+</sup> bone marrow cells of cancer-free patients. In particular, we assessed the expression of selected epithelial (*EPCAM*, *KRT8*, *KRT18*, *KRT19*, *KRT5*, *KRT6a*, *KRT14*), prostate-specific [*KLK3* (PSA)], tumor-specific (*MAGEA2* and *MAGEA4*), common hematopoietic [*PTPRC* (CD45), *CD34*, *CD33*, *CD19*], and erythroid (*GPYC*, *SCL4A1*, *HBA2*) transcripts aiming to separate prostatic epithelial and hematopoietic lineages.

## Materials and Methods

The workflow of this study is given in the Supplementary Fig. S1.

### Patients and bone marrow samples

The local ethics committee of the University of Regensburg (Regensburg, Germany) approved all aspects of the study (ethics vote number 07-079). Between February 2009 and November 2012, bone marrow aspirates from 105 nonmetastatic (M0 stage) prostate cancer patients were collected shortly before radical prostatectomy. In addition, we obtained bone marrow samples of two metastatic (M1 stage) prostate cancer patients. As a control group, 18 bone marrow samples from cancer-free males undergoing trauma or orthopedic surgery were obtained. After obtaining written informed consent, the sampling was performed by aspiration of the left and right iliac crest (12). The samples were transported to the laboratory within 3 hours and subjected to further processing. Baseline patient characteristics are provided in Table 1.

### Processing of bone marrow samples

Detailed description of bone marrow processing is given in Supplementary Materials and Methods. Briefly, upon arrival, the bone marrow sample was twice washed with Hank's balanced salt solution to remove fat and thrombocytes. Next,

cell suspension was centrifuged in density gradient. After centrifugation, the interphase containing mononuclear cells (MNC) were carefully collected and washed with PBS. The number of MNCs and erythrocytes was determined on hemocytometer. To enrich the DCC-containing fraction, the sample was depleted of the majority of hematopoietic cells using negative immunomagnetic selection. This was done by incubating the cell suspension with APC-conjugated antibodies against CD11b, CD33, and CD45. After incubation and washing, the cell suspension was incubated with anti-APC beads and anti-CD235a beads (glycophorin A). After incubation and washing, the cell suspension was run through the 40- $\mu$ m cell sieve, and then run on LS MACS column. The eluate, containing the unlabeled cell fraction, was collected on ice and the cell number determined using a hemocytometer.

### Staining of bone marrow samples and cell lines

Detailed description of staining procedure is given in Supplementary Materials and Methods. On average, two million bone marrow cells were stained with anti-EpCAM-PE (HEA125, Miltenyi Biotec) antibody. Cell lines (WPE-Stem, DU145, PC3, WPE-Int, LNCaP, and VCaP) were grown according to instructions from ATCC. At 80% confluence, the medium was discarded, cell monolayer washed with PBS, and cells fixed using 4% formaldehyde. Further immunostaining steps were performed as described with the addition of nuclear counterstaining using DAPI (Supplementary Fig. S2).

### Screening of bone marrow and isolation of single cells and cell pools

A detailed description of the screening and cell isolation procedures is given in Supplementary Materials and Methods. Each bone marrow sample was manually screened for the presence of EpCAM<sup>+</sup> cells on an inverted fluorescent microscope (Olympus or Zeiss), equipped with micromanipulator (Patchman NP2, Eppendorf) and pump (CellTram, Eppendorf). Single cells with preserved integrity and positively stained for EpCAM were extracted using a glass capillary attached to the micromanipulator. By visual inspection, we ensured that only one cell was in the capillary. The cell was transferred to an empty field with PBS and manually isolated using a micropipette. After isolation of single EpCAM<sup>+</sup> cells from each sample we isolated a pool of approximately 2,000 to 3,000 cells. As a reagent control of cell isolation, 1  $\mu$ L of PBS, in which individual EpCAM<sup>+</sup> cells were isolated, was taken for subsequent whole transcriptome amplification (WTA).

Photographs of the cells before isolation were made in 10 cases, using the software cellSens (Olympus) or AxioVision (Zeiss). Because of the rapid bleaching of PE fluorophore, we abstained from taking photographs of all isolated cells.

Flow-cytometric analysis of bone marrow cells was performed with a LSR II machine equipped with FACS DIVA 5.03 software (BD Biosciences) and data was analyzed with FlowJo 8.8.6 (Treestar).

### Whole transcriptome amplification of single cells

Isolation of mRNA from a single cell, reverse transcription, and global amplification of the first strand cDNA was carried

**Table 1.** Patient baseline characteristics and detection of EpCAM<sup>+</sup> cells.

Clinical characteristics	Overall <i>n</i> (%)	EpCAM <sup>+</sup>		<i>P</i> <sup>a</sup>
		Yes <i>n</i> (%)	No <i>n</i> (%)	
Number of M0-stage prostate cancer patients	105 (100)	65 (62)	40 (38)	0.61
Age, y				
≤60	41 (39)	23 (56)	18 (44)	0.32
>60	64 (61)	42 (66)	22 (34)	
Pre-OP PSA				
≤10 ng/μL	70 (67)	44 (63)	26 (37)	0.77
>10 ng/μL	35 (33)	21 (60)	14 (40)	
Gleason score				
2–6	40 (38)	24 (60)	16 (40)	0.95
7	51 (49)	32 (63)	19 (37)	
8–10	14 (13)	9 (64)	5 (36)	
Pathologic tumor stage				
T2	53 (51)	31 (59)	22 (41)	0.47
T3+4	52 (49)	34 (65)	18 (35)	
Nodal status				
N0	58 (55)	33 (57)	25 (43)	0.32
N+	14 (13)	10 (71)	4 (29)	
n.a.	33 (32)			
Risk classification <sup>b</sup>				
Low and intermediate risk	51 (49)	29 (57)	22 (43)	0.30
High and very high risk	54 (51)	36 (67)	18 (33)	

Abbreviation: n.a., not available.

<sup>a</sup> $\chi^2$  test.<sup>b</sup>According to the European Association of Urology (47).

out as previously described (7, 10). Details are provided in Supplementary Materials and Methods. The quality of amplification was examined by performing endpoint PCR for *ACTB*, *EEF1A1*, and *GAPDH* transcripts (see below and Supplementary Table S1).

#### Whole transcriptome amplification controls

From all samples, WTA controls were generated by aspirating 1 μL buffer without a cell from the slide after DCC isolation and equally subjected to WTA. After WTA, we controlled for the presence of *ACTB*, *EEF1A1*, and *GAPDH* transcripts. Only when WTA controls were negative, cells isolated during the same experiment were included into transcriptome analysis. In 2 of 125 (2%) bone marrow samples, the WTA control was contaminated. The WTA samples of these two patients were excluded from genome and transcriptome analysis.

#### Whole genome amplification and comparative genomic hybridization

Single-cell genomic DNA collected during WTA procedure was precipitated and subjected to whole genome amplification (WGA), using the Ampli1 WGA Kit (Silicon Biosystems), as described previously (10, 13). Details are provided in Supplementary Materials and Methods. CGH on metaphases (cCGH, *n* = 24 samples) and Agilent microarrays (aCGH, *n* = 18 samples) was

performed as described previously (10, 13, 14). Array files are deposited in GEO database under accession number GSE59631.

#### Expression analysis of selected transcripts by endpoint PCR in patient samples

All WTA products were tested by endpoint PCR using conditions outlined in Supplementary Materials and Methods. The list of primers is given in Supplementary Table S1. To establish the correct identity of each amplified fragment, we performed restriction digestion, using the conditions outlined in Supplementary Materials and Methods. Only after a positive result after digestion, the sample was declared positive for a given transcript.

#### Statistical analyses

Statistical analyses were performed using SPSS v22 (IBM). Hierarchical clustering was performed using the software tool Cluster (15).

#### Results

##### Detection of EpCAM<sup>+</sup> cells in bone marrow samples and relation to clinical variables

We investigated the presence of putative DCCs in bone marrow samples of 105 M0-stage prostate cancer patients undergoing radical prostatectomy by staining for EpCAM. In



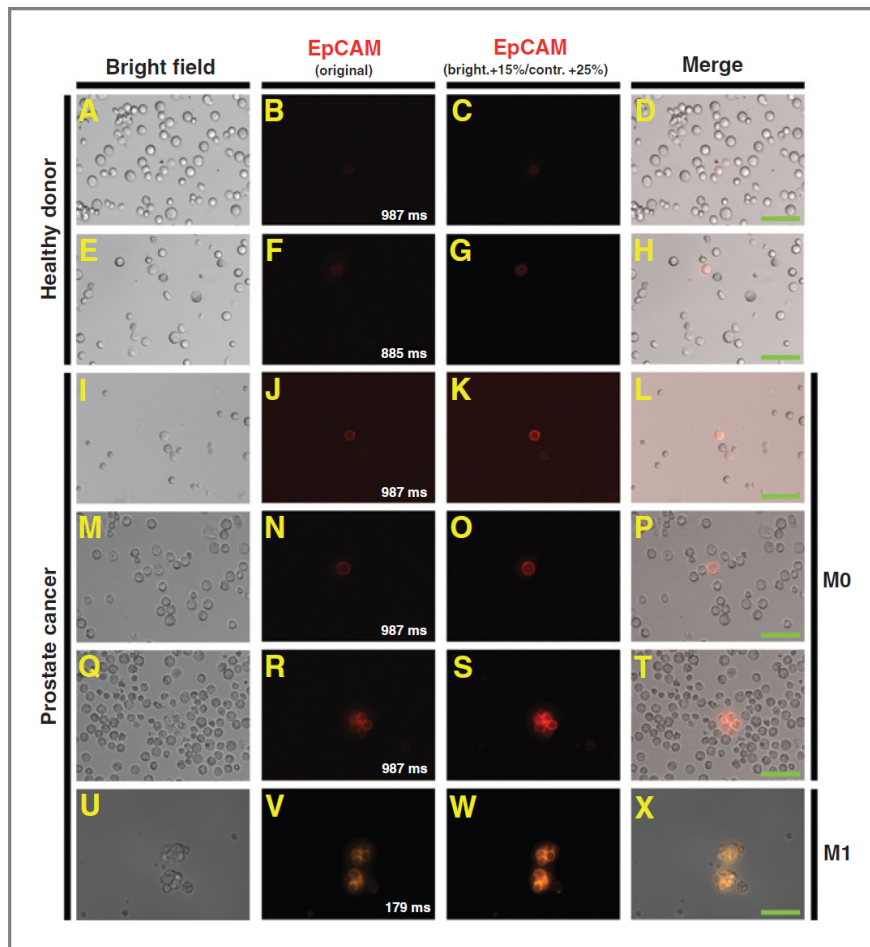


Figure 1. Intensity of EpCAM staining of cells isolated from patients with prostate cancer and controls. Photographs of the EpCAM<sup>+</sup> cells isolated from two healthy donors (A–D and E–H) and three patients with prostate cancer in clinical stage M0 (I–L and M–T) and M1 (U–X). Photographs (A–D, E–H, I–L, and M–P) show individual EpCAM<sup>+</sup> cells, and photographs Q–T and U–X show clusters of EpCAM<sup>+</sup> cells. Images A, E, I, M, Q, and U depict bright-field photographs. Images B, C, F, G, J, K, N, O, R, S, V, and W depict photographs of cells taken under fluorescent microscopy (Cy3 channel). Numbers in the bottom right corner of the images B, F, J, N, R, and V ["EpCAM (original)"] represent the exposure time used for taking of the photograph. Note that photographs U–X were taken on a different microscope (Zeiss), compared with photographs A–T (Olympus), hence different exposure times. Images C, G, K, O, S, and W ["EpCAM (bright.+15%/contr.+25%)"] were manipulated in a way that brightness and contrast of the original image were increased, as indicated, using image processing software to demonstrate the different staining intensity of EpCAM<sup>+</sup> cells of controls and patients with cancer. Images D, H, L, P, T, and X are the result of the merging of corresponding original bright-field and fluorescent photographs. Scale bar, 50  $\mu$ m.

addition, we stained bone marrow samples from two M1-stage patients with prostate cancer and 18 male control patients without cancer (Supplementary Fig. S1). To enrich the epithelial target population, we first depleted bone marrow samples of CD45<sup>+</sup>, CD33<sup>+</sup>, CD11b<sup>+</sup>, and glycophorin A<sup>+</sup> cells. On average, 86% of MNCs were removed by immunomagnetic sorting (Supplementary Fig. S3). Then, we used direct immunofluorescence to detect EpCAM<sup>+</sup> cells. Screening on average  $2 \times 10^6$  cells per bone marrow sample (median  $2 \times 10^6$ , range  $0.02$ – $6.8 \times 10^6$  cells; Supplementary Fig. S3) revealed that 65 (62%) of M0-stage patients, as well as 10 (56%) of control patients had EpCAM<sup>+</sup> cells in the bone marrow ( $P = 0.79$ , Fisher exact test, two-tailed; Supplementary Table S2). Both analyzed M1-stage patients also harbored EpCAM<sup>+</sup> cells in the bone marrow.

Clinical characteristics of the M0-stage cohort before radical prostatectomy are shown in Table 1. There was no significant association between the presence of EpCAM<sup>+</sup> cells in the bone marrow of M0-stage prostate cancer patients and any clinical risk factor such as Gleason score or preoperative PSA (Table 1). Likewise, grading (G), resection margin (R), lymphatic invasion (L), or vein invasion (V)

and perineural infiltration (PnI) were not associated with detection of EpCAM<sup>+</sup> cells (data not shown).

#### EpCAM staining of bone marrow samples from patients and controls

EpCAM<sup>+</sup> cells from patients with prostate cancer (Fig. 1I–X) usually displayed stronger staining intensity compared with cells from controls (Fig. 1A–H). However, solely based on staining intensity, size, or morphology, one would not be able to reliably identify true DCCs. Because of rapid fading of phycoerythrin, we abstained from taking pictures of each cell or to perform real-time measurements of fluorescence intensity. Still, comparison of available photographs of EpCAM<sup>+</sup> cells from patients and controls, with EpCAM-stained prostate epithelial cell lines (Supplementary Fig. S4) suggests that EpCAM<sup>+</sup> cells from bone marrow of controls and M0-stage patients display a staining pattern similar to cell lines with low levels of EpCAM expression such as PC3 or DU145 (Supplementary Fig. S4E–S4P). To corroborate these observations, we used a bone marrow sample of a cancer-free patient, depleted it of CD45<sup>+</sup>, CD33<sup>+</sup>, CD11b<sup>+</sup>, and glycophorin A<sup>+</sup> cells, and

divided the sample in halves. We spiked one half with 2% of EpCAM<sup>+</sup> DU145 prostate cancer cell line cells. Then, spiked and nonspiked samples were again divided and stained using a EpCAM-PE antibody or isotype-PE antibody. No fluorescent signal was observed in samples stained with isotype-PE antibody (Supplementary Fig. S5A–S5H), whereas weakly EpCAM<sup>+</sup> cells could be seen in both samples stained with EpCAM-PE antibody (Supplementary Fig. S5I–S5L). As expected, these cells displayed slightly weaker EpCAM positivity compared with some of DU145 cells (Supplementary Fig. S5M–S5P). Still, only the top 17% of the spiked-in DU145 population (i.e., 0.34% of all EpCAM<sup>+</sup> cells) could be identified according to their higher EpCAM positivity by flow-cytometric analysis (Supplementary Fig. S5Q). This shows that the intensity of the EpCAM staining can rarely be used to discriminate EpCAM<sup>+</sup> true DCCs from EpCAM<sup>+</sup> bone marrow cells.

#### Isolation of EpCAM<sup>+</sup> single cells and whole genome and transcriptome amplification

On average, we isolated four individual EpCAM<sup>+</sup> cells from each positive bone marrow sample (Supplementary Fig. S3). From 65 EpCAM<sup>+</sup> bone marrow samples of M0-stage prostate cancer patients, we isolated 220 EpCAM<sup>+</sup> single cells (examples shown in Fig. 1I–P) and five clusters of two to five cells (examples shown in Fig. 1Q–T), and twelve EpCAM<sup>+</sup> single cells and four clusters of two to five cells from two M1-stage prostate cancer patients (Fig. 1U–X). Fifty-one individual EpCAM<sup>+</sup> cells were collected from ten control patients (Fig. 1A–H; Supplementary Table S2). Bone marrow of controls never contained EpCAM<sup>+</sup> cell clusters. After isolation of EpCAM<sup>+</sup> cells, we isolated a pool of approximately 2,000 to 3,000 cells from each bone marrow sample.

Next, we extracted and separated total genomic DNA (gDNA) and mRNA from each single cell, cell cluster, or cell pool. Total mRNA was reversely transcribed and subjected to WTA, while gDNA was subjected to WGA. For analysis of single-cell transcriptomes, only high-quality samples were selected on the basis of our quality control (QC) criteria (i.e.,

positive for the three QC PCR transcripts *ACTB*, *EEF1A1*, and *GAPDH*, see Materials and Methods).

The three groups (M0, M1, controls) did not differ with regard to cDNA quality ( $P = 0.75$ , Pearson  $\chi^2$ ; Supplementary Table S2 and Supplementary Fig. S7A). After WTA and QC, we selected 119 EpCAM<sup>+</sup> single cells and five clusters of EpCAM<sup>+</sup> cells (55% of samples) from M0-stage patients. From controls, we selected 28 (55%) of EpCAM<sup>+</sup> single cells, and from M1-stage patients seven EpCAM<sup>+</sup> single cells and three clusters of EpCAM<sup>+</sup> cells (62% of samples; Supplementary Table S2 and Supplementary Fig. S7A). In addition, high-quality WTA products of bone marrow cell pools from ten randomly selected M0-stage patients, one M1-stage patient, and ten controls were chosen for further analysis.

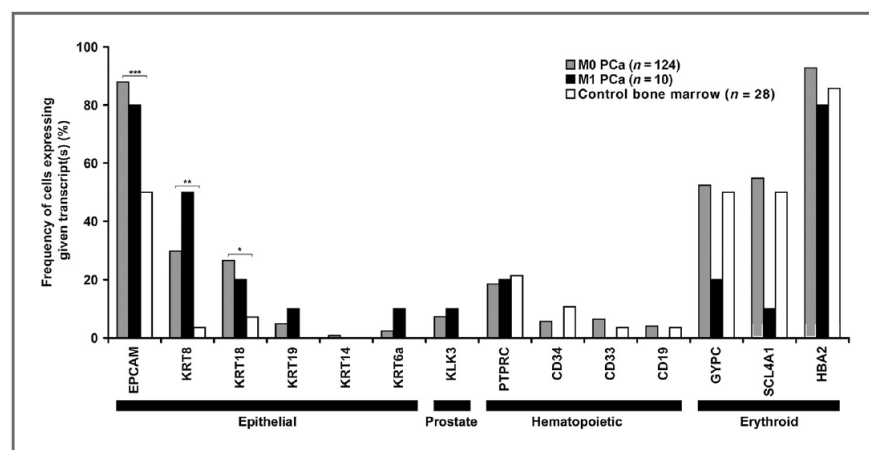
#### Expression profiling for identification of a DCC signature in M0-stage patients

We started the search for a DCC-specific signature by expression analysis of selected genes in bone marrow cell pools of prostate cancer patients and controls. We could not detect significantly different frequencies for epithelial-, prostate-, tumor-associated, and erythroid transcript expression between the analyzed cell pools (Supplementary Table S3; Supplementary Figs. S6 and S8D–S8F). Surprisingly, among common hematopoietic transcripts, only *CD19* was expressed in a significantly higher number of samples from M0-stage patients as compared with pools from healthy controls ( $P < 0.05$ , Fisher exact test).

As the analysis of bone marrow cell pools revealed no transcript specific for cancer patients or controls, we focused on expression profiling of selected transcripts in EpCAM<sup>+</sup> single cells. We first analyzed the expression of *EPCAM* and cytokeratins transcripts. The number of EpCAM<sup>+</sup> single cells expressing the *EPCAM* transcript was significantly higher among cells from M0-stage patients (88%), than among cells from controls (50%;  $P < 0.001$ , Pearson  $\chi^2$ ; Fig. 2 and Supplementary Fig. S7B; Supplementary Table S3).

Forty-two percent of cells from M0-stage patients expressed at least one *KRT* transcript (Supplementary Fig. S7B),

**Figure 2.** Gene expression of EpCAM<sup>+</sup> cells isolated from prostate cancer (PCa) and control patients. The identities of transcripts are on the x-axis. Results for *KRT5*, *MAGEA2* and *MAGEA4* are not shown, as the transcripts were not detected in any sample. Exact numbers and percentages are given in Supplementary Table S3. M0- (gray bars) and M1-stage (black bars) prostate cancer patients and control patients (white bars). Asterisks, the level of significance: \*,  $P < 0.05$ ; \*\*,  $P < 0.01$ ; \*\*\*,  $P < 0.001$ ; n, the number of analyzed single cells.



which was significantly different to cells from controls, where only 7% of cells expressed *KRT* transcripts ( $P < 0.01$ ,  $\chi^2$ ). The difference between patients with cancer and controls was even more pronounced, when we analyzed *KRT* expression in *EPCAM*-transcript-positive cells. We found that cells from M0-stage patients more frequently coexpressed *EPCAM* and *KRT* transcripts, compared with cells from controls ( $P < 0.002$ , Fisher exact test; Supplementary Table S4). Among *KRT*-expressing cells from M0-stage patients, the most frequently detected *KRT* transcripts were *KRT8* and *KRT18* (Fig. 2 and Supplementary Fig. S7C; Supplementary Table S3). These transcripts were detected significantly more often in cells from M0-stage patients than in controls (*KRT8*, 30% vs. 4%,  $P < 0.01$ ; *KRT18*, 27% vs. 7%,  $P < 0.05$ ; Pearson  $\chi^2$ ; Supplementary Table S3). *KRT19*, *KRT14*, and *KRT6a* transcripts were found in less than 5% of cells from M0-stage patients, and were absent in cells from controls. No *KRT5* transcripts were detected in cells of either group (Supplementary Table S3 and Supplementary Fig. S7C). In total, among cells from M0-stage patients, 93% of cells expressed at least one of the tested epithelial transcripts, compared with 57% cells from controls ( $P < 0.001$ , Fisher exact test). This distribution closely resembled the frequency of *EPCAM* expression observed in these two groups of cells (Supplementary Fig. S7B).

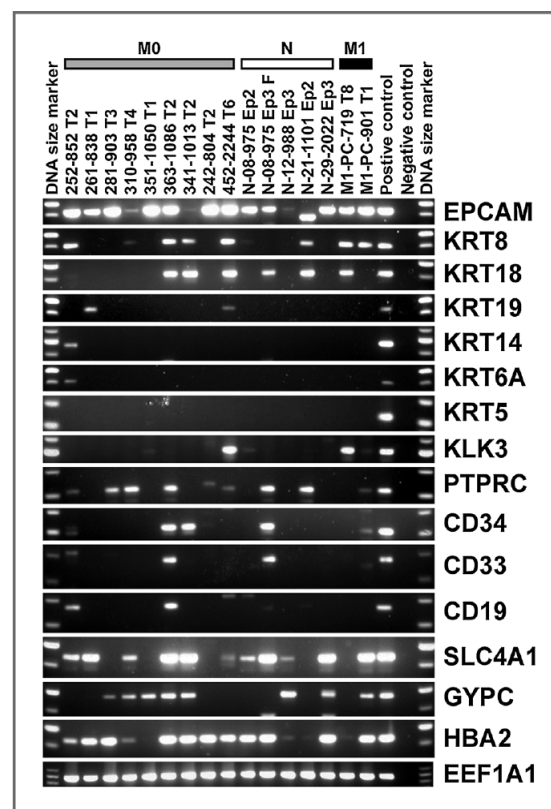
Next, we analyzed prostate- and tumor-specific transcripts. *KLK3* (PSA) transcript was found only in cells isolated from one M0-stage patient, whose bone marrow cell pool also expressed *KLK3*. Overall, this transcript was detected in 9 (7%) M0-stage cells and was absent in cells from controls ( $P = 0.143$ , Fisher exact test; Fig. 2; Supplementary Table S3). *MAGEA2* and *MAGEA4* transcripts could not be found in any of the *EpCAM*<sup>+</sup> single cells.

Given the rare detection of histogenetic or cancer-associated epithelial transcripts, we sought to determine whether hematopoietic transcripts are useful to differentiate between *EpCAM*<sup>+</sup> cells from cancer versus control patients. For this, we analyzed the expression of selected hematopoietic transcripts. To our surprise, among cells from M0-stage cancer patients, the *PTPRC* transcript (CD45) was most frequently detected (19% of cells; Fig. 2; Supplementary Table S3). Overall, 23% of *EpCAM*<sup>+</sup> single cells from M0-stage prostate cancer patients expressed at least one of the analyzed hematopoietic transcripts (*PTPRC*, *CD34*, *CD33*, or *CD19*). Similarly, 29% of *EpCAM*<sup>+</sup> single cells from control patients expressed at least one of mentioned hematopoietic transcripts, of which *PTPRC* was also most frequent (21% of cells; Fig. 2; Supplementary Table S3). In single cells, the frequencies of hematopoietic transcript expression did not differ significantly ( $P = 0.5$ ; Pearson  $\chi^2$ ). We then analyzed the expression of selected erythroid lineage-specific transcripts, *GYPC* (CD236, glycophorin C), *SCL4A1* (CD233, band 3 protein), and *HBA2* (hemoglobin  $\alpha 2$ ). Frequencies of erythroid transcript-expressing cells between M0-stage patients and controls were almost identical ( $P = 0.5$ , Pearson  $\chi^2$ ). Here, the most frequently detected transcript was *HBA2* and it was detected in 93% of cells from M0-stage patients and in 86% of cells from controls (Fig. 2; Supplementary

Table S3). Overall, 96% of cells from M0-stage patients expressed at least one erythroid transcript, while the frequency of such cells among control cells was 93%.

This unexpected observation prompted us to control whether extracellular erythroid transcripts may contaminate the single-cell transcriptomes by mRNA carryover or release during sample processing. Therefore, we checked the WTA controls (see Materials and Methods section for details) of 40 patients for the presence of *HBA2* transcript. As all 40 samples were negative (Supplementary Fig. S9) we note a striking difference to the cell-containing samples of M0-stage prostate cancer patients ( $P < 0.0001$ , Pearson  $\chi^2$ ).

Overall, analysis of epithelial, hematopoietic, and erythroid transcripts in *EpCAM*<sup>+</sup> single cells from cancer patients and controls revealed unexpected transcriptomes (Fig. 3 and Supplementary Fig. S8A and S8B).



**Figure 3.** Transcript profiling by PCR of WTA cDNA of selected *EpCAM*<sup>+</sup> single cells or small clusters of *EpCAM*<sup>+</sup> cells from patients with prostate cancer and controls. The photograph of each agarose gel was cropped to contain the band of interest and at least two nearest DNA size marker bands. The sizes of fragments are given in Supplementary Table S1 and Supplementary Fig. S11. Identifiers of the samples are given above the wells, and the identity of analyzed transcript is given next to each cropped gel. Only those samples that contain the band of correct size were considered positive. The brightness and contrast of each individual cropped image were manipulated by using the "Auto contrast" tool of Adobe Photoshop.



### Expression profiling of cells from M1-stage patients

For the majority of analyzed genes, EpCAM<sup>+</sup> single cells or cell clusters from M1-stage cancer patients displayed similar frequencies of transcript-positive cells similar to those from M0-stage patients (Fig. 2, Fig. 3, and Supplementary Fig. S7B; Supplementary Table S3). A detailed overview of coexpression of individual transcripts is given in Supplementary Fig. S8C. As we obtained only two bone marrow samples and selected seven EpCAM<sup>+</sup> single cells and three clusters for further analysis, a statistical comparison with the other groups was not performed.

### Patterns of gene expression profiles and phenotypic heterogeneity

None of the analyzed transcripts could reliably discriminate between EpCAM<sup>+</sup> cells from patients and controls. We therefore asked whether a combination of analyzed transcripts could help us to identify true DCCs. When we performed hierarchical cluster analysis, all EpCAM<sup>+</sup> cells (from M0 and M1-stage patients and controls) were evenly scattered among dendrogram branches (Fig. 4A). The analysis of cell pools revealed two major clusters; however, pools from patients with prostate cancer and controls were evenly distributed between the clusters (Fig. 4B).

From 35 (54%) EpCAM<sup>+</sup> M0-stage bone marrow samples, we isolated two or more cells. Among these 35 patients, cells from 14 patients (40%) clustered together. However, 12 of 14 patients had at least one cell that did not cluster with others, whereas only in the case of two patients, all cells clustered together. Therefore, the great majority of samples with two or more isolated cells contained cells with heterogeneous expression of analyzed transcripts. In controls, we found 5 of 10 samples with more than one cell. Here, 2 of 5 (40%) clustered together, with at least one cell that did not cluster with others ( $P = 1$  for comparison with M0-stage patients; Fisher exact test). Cells from one M1-stage patients clustered together with several cells scattered within the dendrogram (Fig. 4A).

### Analysis of genomic aberrations

The presence of hematopoietic transcripts in EpCAM<sup>+</sup> single cells from patients with prostate cancer casted doubts whether these are indeed cancer cells, although many of them coexpressed epithelial transcripts. This prompted us to analyze selected cells for the presence of genomic aberrations using CGH. While the absence of genomic aberrations at the given resolution does not prove that the analyzed cell is not a cancer cell (16), the presence of genomic aberrations confirms the malignant origin of the cell.

To demonstrate the reliability of our combined genome and transcriptome analysis, we first compared the profile of a single VCaP cell with the profile of VCaP unamplified genomic DNA and found that the majority of genomic aberrations could be retrieved at single cell genome level (Supplementary Fig. S10A). In contrast, two peripheral blood leukocytes from a healthy volunteer displayed normal karyotypes (Supplementary Fig. S10B).

Next, we analyzed the genome of four EpCAM<sup>+</sup> cells from controls, and 34 EpCAM<sup>+</sup> cells isolated from the bone marrow

of M0- and M1-stage patients using CGH. No aberrations could be detected in the genomes of four EpCAM<sup>+</sup> cells from controls, (Supplementary Fig. S10C), whereas 13 of 34 cells (38%) from patients had genomic aberrations. Strikingly, we found cells expressing hematopoietic (e.g., CD45) or erythroid (e.g., *HBA2*) transcripts, and harboring genomic aberrations, both in M0- and M1-stage patients (Fig. 5 and Supplementary Fig. S10E). These data demonstrate that DCCs can express hematopoietic and erythroid transcripts.

Finally, we compared the 17-gene signature of EpCAM<sup>+</sup> cells from patients with prostate cancer (M0 and M1 stage) versus EpCAM<sup>+</sup> cells selected for genomic aberrations. Overall, the expression patterns did not change, although we noted an increased frequency of cells expressing epithelial transcripts (*KRT8*, *KRT18*, *KRT19*, *KRT6a*, *EPCAM*) among cells harboring genomic aberrations. Furthermore, erythroid marker expression was clearly detectable in these cells similar to the whole cohort (e.g., *HBA2* was detected in 77% vs. 92% and CD45 in 15% vs. 18%, respectively). A potentially relevant change was observed for PSA (*KLK3*). Here, cells harboring genomic aberrations more often expressed the transcript than the whole group of EpCAM<sup>+</sup> cells from patients with prostate cancer ( $P = 0.02$ ; Fisher exact test; Fig. 6).

### Discussion

In this study, we performed targeted expression profiling of EpCAM<sup>+</sup> single cells isolated from the bone marrow of patients with nonmetastatic (M0 stage) prostate cancer and demonstrate, contrary to our expectation, that a 17-gene mRNA signature based on selected histogenetic markers is insufficient to reliably identify DCCs. However, by combining genome and transcriptome analysis, we could show that prostate cancer-DCCs mimic the transcriptomes of bone marrow cells and harbor hematopoietic and particularly erythroid transcripts. These findings suggest a high phenotypic plasticity of cancer cells clearly surpassing the classical concept of epithelial-mesenchymal transition (17).

We started by analyzing cell pools of CD45-, CD33-, CD11b-, and glycophorin A-depleted bone marrow cells from patients with prostate cancer and controls, reasoning that one or several markers should be able to classify cancer-derived samples correctly. As this was not the case, we focussed on individually isolated single EpCAM-positive cells expecting that this population contained *bona fide* DCCs with a defined transcriptional profile.

EpCAM protein expression appeared to be stronger in cells from patients with prostate cancer than controls. However, the percentage of positive samples was comparable between groups and reliable identification of DCCs based on EpCAM staining alone was not possible (10, 11, 18). Of note, EpCAM expression by putative DCCs in bone marrow was rather low compared with the typical appearance of most prostate cancer cell lines (19, 20). Consequently, we investigated whether our marker panel, comprising epithelial (*EPCAM*, *KRT8*, *KRT18*, *KRT19*, *KRT14*, *KRT6a*, *KRT5*), the prostate-specific *KLK3* (PSA), tumor-associated (*MAGEA2*, *MAGEA4*) and hematopoietic transcripts [*PTPRC* (CD45), *CD33*, *CD34*, *CD19*, *GYPC*,

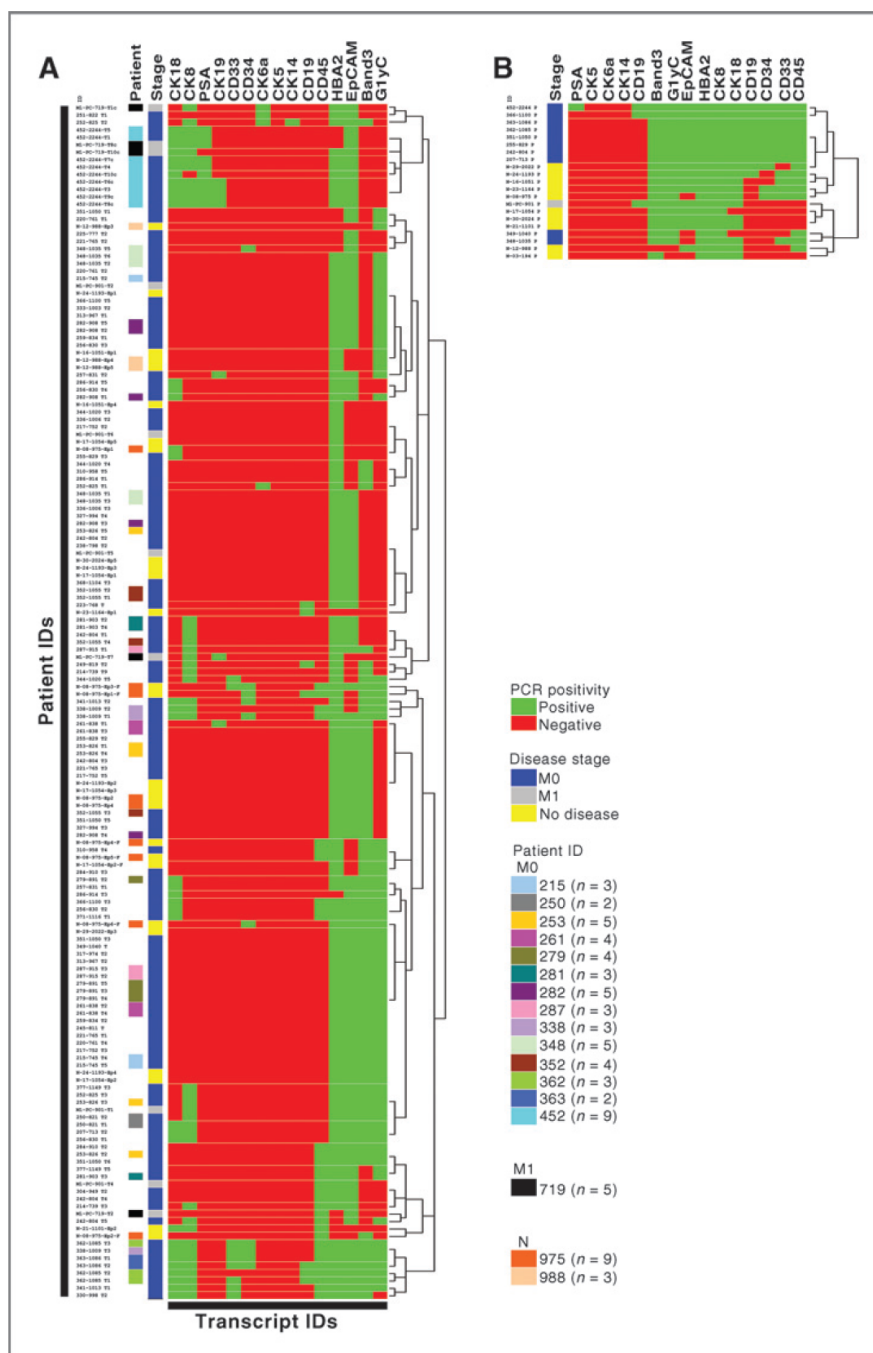


Figure 4. Hierarchical cluster analysis of gene expression. Single cells (A) and pools from M0- and M1-stage patients (B), and healthy controls. Green color in heat maps denotes that the marker is detected by PCR in the sample, while red color indicates that the sample was negative. The sample type [M0, M1 stage, or N (healthy)] is color coded. Groups of cells that clustered together and were isolated from the same patient are also color coded. Related Supplementary Fig. S8 displays all results of expression profiling grouped by sample type (single cell or cell pool) and source (patients with cancer or controls).

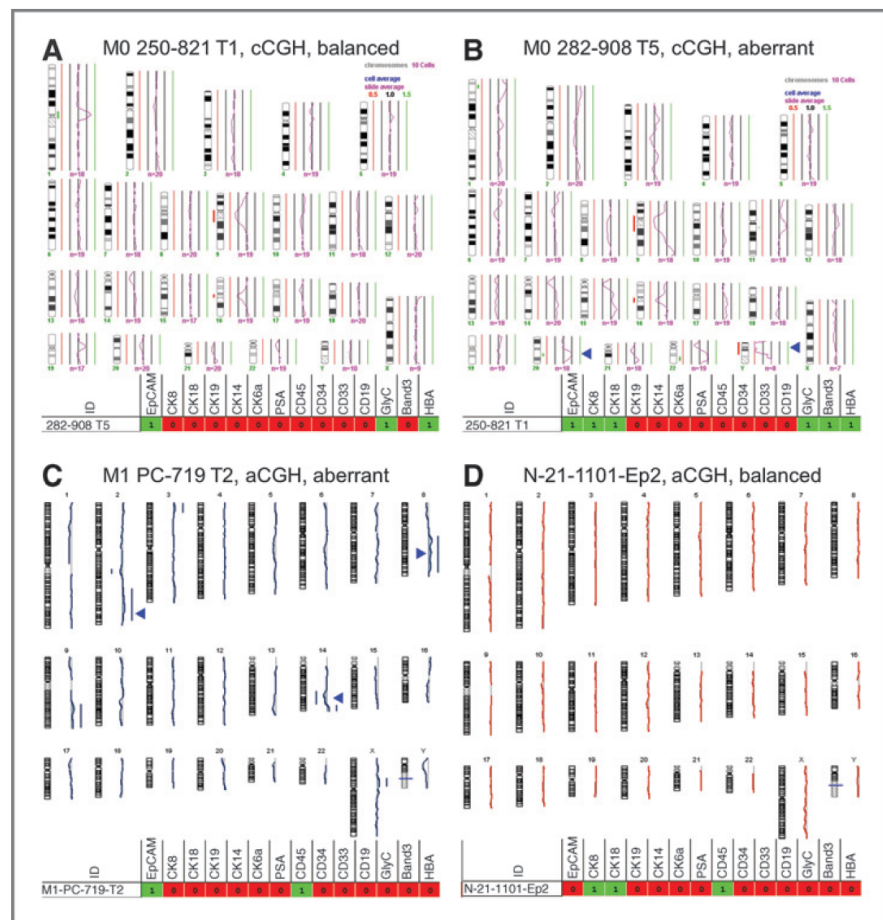
*SCL4A1* (band 3), *HBA2*], would discriminate EpCAM<sup>+</sup> cells from patients with prostate cancer and controls.

For all analyzed transcripts, we established single cell sensitivity. However, none of the transcripts enabled a clear DCC identification, either because very few DCCs were positive or because control cells expressed a given transcript at comparable frequency. Similarly, we identified no expres-

sion pattern characteristic for DCCs or EpCAM-positive cells from controls.

These findings, while unexpected, corroborate previous circumstantial observations. First, EpCAM transcript detection correlated with the fluorescence staining intensity. More than 88% of the subjectively brighter stained EpCAM<sup>+</sup> single cells from patients with prostate cancer harbored the EpCAM

Figure 5. Combined genome and transcriptome analysis of single EpCAM<sup>+</sup> cells. A and B, single-cell CGH profiles of two M0-stage and (C) one M1-stage and one control patient (D). Some of the detected aberrations are indicated by arrowheads. Centromeric regions known to generate false positive signals are excluded from analysis. Results of the transcript profiling of each cell is given below the CGH profile (green and 1 indicate that the transcript was detected in the sample, whereas red and 0 indicate that the transcript was not detected). The remaining CGH profiles and corresponding expression profiles are given in Supplementary Fig. S10. cCGH, CGH on metaphase spreads; aCGH, CGH on DNA microarrays.



transcript, whereas we detected EpCAM transcripts in only 50% of weakly stained EpCAM<sup>+</sup> control cells, suggesting a high technical reliability of our study. EpCAM, a highly glycosylated protein, is known to have a long half-life (21), which may explain low transcription in control cells. Second, significantly more EpCAM<sup>+</sup> cells from patients with prostate cancer coexpressed epithelial transcripts than EpCAM<sup>+</sup> cells from controls (93% vs. 57%). This coexpression included particularly *KRT8* and *KRT18*, whose protein products are recognized by anti-CK antibodies, such as A45-B/B3, commonly used to detect DCCs (12), whereas other *KRT* transcripts were essentially absent from EpCAM<sup>+</sup> cells of patients with prostate cancer. Thus, EpCAM<sup>+</sup> cells from patients with prostate cancer may comprise cells of malignant epithelial origin, i.e., DCCs, but these DCCs do not express cytokeratins typical for basal, stem-like cells, such as *KRT5*, *KRT6a*, and *KRT14* (22–25). Third, the transcript for the prostate-specific differentiation marker PSA (*KLK3*) or cancer-associated (*MAGEA*; ref. 26) transcripts were hardly or not detected in EpCAM<sup>+</sup> cells of patients with prostate cancer. The absence of *MAGEA* transcripts is in line with low detection rates in analyses of bulk bone marrow (27), and the rare finding of *KLK3* expression with similar studies on CK<sup>+</sup> cells (28, 29).

However, others have found higher frequency of EpCAM<sup>+</sup> DCCs expressing PSA (30, 31), suggesting that EpCAM<sup>+</sup> DCCs may either be undifferentiated or that PSA expression is linked to the formation of cell clusters and micrometastases.

Even analysis of hematopoietic markers did not help to differentiate DCCs and the native EpCAM<sup>+</sup> bone marrow cell population. For single cells, we could not find significant differences in the frequency of *PTPRC* (CD45), *CD19*, *CD33*, *CD34* transcript-containing cells between M0-stage and control patients. As published data suggested that cells of the early erythroid lineage in the bone marrow could express EpCAM (32, 33), we tested the presence of erythroid-specific transcripts, namely *HBA2* (34), *SCL4A1* (35), and *GYPC* (36). Expression was abundant in both groups of cells, with EpCAM<sup>+</sup> cells from controls and patients with prostate cancer expressing at least one of the three transcripts in 93% and 96%, respectively. In both groups, *HBA2* was expressed most frequently, which was also detected in 80% of EpCAM<sup>+</sup> cells from M1-stage patients.

Despite the fact that EpCAM<sup>+</sup> single cells from M0- and M1-stage patients are clearly different from EpCAM<sup>+</sup> cells from controls, the findings substantially reduce our ability to correctly identify an individual cell as DCC from this 17-gene



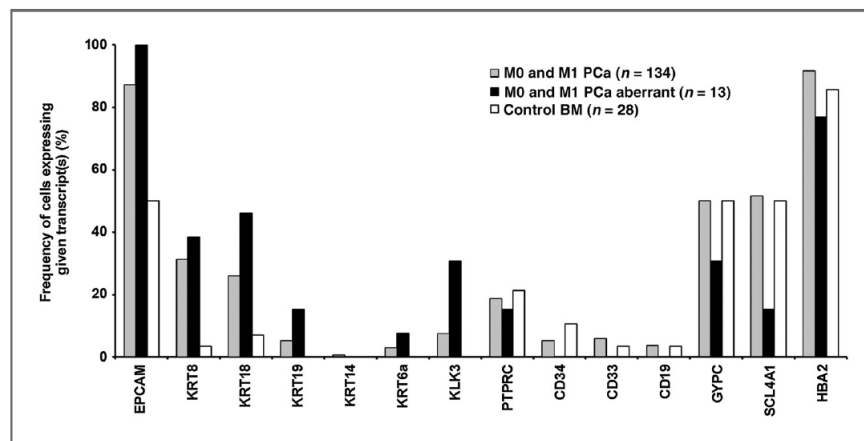


Figure 6. Gene expression of DCCs from patients with prostate cancer (PCa). Cells with proven malignant origin display similar profiles as the complete group of EpCAM<sup>+</sup> cells from patients with prostate cancer. Transcripts are identified by gene symbols. Results for *KRT5*, *MAGEA2*, and *MAGEA4* are not shown, as the transcripts were not detected in any sample. EpCAM<sup>+</sup> cells from M0- and M1-stage (gray bars) prostate cancer patients, EpCAM<sup>+</sup> cells with genomic aberrations from M0- and M1-stage (black bars) prostate cancer patients, and EpCAM<sup>+</sup> cells from control patients (white bars). n, the number of analyzed single cells.

transcriptome. We therefore first excluded the possibility that the *HBA2* detection in our samples resulted from carryover of extracellular mRNA. We then sought for independent validation that EpCAM<sup>+</sup> cells of patients with prostate cancer are truly cancer cells and performed combined genome and transcriptome analysis of selected single cells. Here, we noted that cells of proven malignant origin more frequently express cytokeratin markers as compared with the total group of EpCAM<sup>+</sup> cells, suggesting that EpCAM<sup>+</sup> normal cells may have diluted true DCCs. In fact, expression of *KLK3* (PSA) was significantly enriched among aberrant EpCAM<sup>+</sup> cells and apparently is the only marker that correctly identifies prostate cancer-DCCs. However, this finding must be validated as *KLK3*-expressing cells were isolated from only two patients. Still, many genomically proven prostate cancer-DCCs did not express *KLK3* but a similar pattern as the remaining EpCAM<sup>+</sup> cells from patients with prostate cancer. Thus, genomically aberrant EpCAM<sup>+</sup> cells contain transcripts of common hematopoietic markers, such as CD45, as well as erythroid markers, demonstrating an unexpected transcriptomic plasticity of both M0- and M1-stage prostate cancer DCCs.

Our findings have important implications for the study of metastasis. First, an educated guess for CTC or DCC-specific markers seems highly unreliable. So far, many studies assume histogenetic specificity (37) that we could not confirm for transcriptomes of bone marrow-derived DCCs. Likewise attempts to reconstruct cancer cell transcriptomes from unpurified CTCs by transcript subtraction need to be viewed with caution (38). Our findings are consistent with previous observations, such as a transcriptome study of 15 single prostate CTCs from one M1-stage patient revealing low-level expression of *EPCAM*, *PTPRC* (CD45), and *CD34* (39), expression of hemoglobins in breast cancer epithelial cells (40, 41) and ovarian cancer epithelial cells transdifferentiating into erythroid cells expressing hemoglobins (42). The observed expression of hematopoietic and erythroid transcripts by bone marrow-DCCs may reflect adaptation to the specific environment. Of note, it was shown that metastatic prostate cells compete with hematopoietic stem cells (HSC) for the HSC

niche (43). If DCCs indeed reside in the HSC niche, it is possible that crosstalk between DCCs and bone marrow cells, mediated by cell-cell interaction or secreted factors, leads to expression of hematopoietic transcripts in DCCs. Furthermore, the transcriptomes of DCCs could be altered by the uptake of exosomes released from resident bone marrow cells (44) or our observations could be the consequence of hypoxic/oxidative stress imposed on DCCs within the bone marrow microenvironment (45, 46). While additional studies are needed to identify the underlying causes, the observed altered transcriptomes are likely to result in altered phenotypes and cell function, which may be highly relevant for our attempts to tailor cancer-specific therapies.

In summary, until a comprehensive analysis of genes expressed by EpCAM<sup>+</sup> cells of patients with prostate cancer versus EpCAM<sup>+</sup> cells of controls is performed, we will struggle to identify new markers specific for DCCs or confounding cells of nonepithelial lineage. Detection of genomic aberrations currently seems to be the most reliable way to confirm cancer cell identity.

#### Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

#### Authors' Contributions

**Conception and design:** M. Gužvić, B. Braun, C.A. Klein

**Development of methodology:** M. Gužvić, B. Braun, R. Ganzer, M. Nerlich

**Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.):** M. Gužvić, B. Braun, R. Ganzer, M. Burger, M. Nerlich, S. Winkler, M. Werner-Klein, Z.T. Czyż, B. Polzer

**Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis):** M. Gužvić, B. Braun, Z.T. Czyż, B. Polzer

**Writing, review, and/or revision of the manuscript:** M. Gužvić, B. Braun, S. Winkler, M. Werner-Klein, Z.T. Czyż, C.A. Klein

**Study supervision:** C.A. Klein

#### Acknowledgments

The authors thank Isabell Blochberger, Sandra Grunewald, Manfred Meyer, Irene Nebeja, Katharina Schneider, and Thomas Schamberger for excellent technical help, and to Elisabeth Doblinger, Gundula Haunschild, Nina Patwary, and Giancarlo Feliciello for help with collecting the data. The authors also thank Tumorzentrum Regensburg e.V. for help in collecting patient data and Verena Huppert, Ana Grujović, and Norbert Braun for critical reading of the manuscript.

## Grant Support

This work was supported by the awards from Prostate Cancer Foundation and Dr. Josef Steiner Cancer Research Foundation Award, and grants from EU (FP6 STREP Research Consortium Migrating Cancer Stem Cells), Deutsche Krebshilfe (Research Consortium Tumour Stem Cells), and Bavarian Genome Research Network (Baygene) to C.A. Klein.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received April 1, 2014; revised October 7, 2014; accepted October 7, 2014; published OnlineFirst October 15, 2014.

## References

- Brabletz T, Lyden D, Steeg PS, Werb Z. Roadblocks to translational advances on metastasis research. *Nat Med* 2013;19:1104–9.
- Allard WJ, Matera J, Miller MC, Repollet M, Connelly MC, Rao C, et al. Tumor cells circulate in the peripheral blood of all major carcinomas but not in healthy subjects or patients with nonmalignant diseases. *Clin Cancer Res* 2004;10:6897–904.
- Braun S, Vogl FD, Naume B, Janni W, Osborne MP, Coombes RC, et al. A pooled analysis of bone marrow micrometastasis in breast cancer. *N Engl J Med* 2005;353:793–802.
- Stoecklein NH, Hosch SB, Bezler M, Stern F, Hartmann CH, Vay C, et al. Direct genetic analysis of single disseminated cancer cells for prediction of outcome and therapy selection in esophageal cancer. *Cancer Cell* 2008;13:441–53.
- Weckermann D, Polzer B, Ragg T, Blana A, Schlimok G, Arnholdt H, et al. Perioperative activation of disseminated tumor cells in bone marrow of patients with prostate cancer. *J Clin Oncol* 2009;27:1549–56.
- Klein CA, Blankenstein TJ, Schmidt-Kittler O, Petronio M, Polzer B, Stoecklein NH, et al. Genetic heterogeneity of single disseminated tumour cells in minimal residual cancer. *Lancet* 2002;360:683–9.
- Hartmann CH, Klein CA. Gene expression profiling of single cells on large-scale oligonucleotide arrays. *Nucleic Acids Res* 2006;34:e143.
- Morgan TM, Lange PH, Porter MP, Lin DW, Ellis WJ, Gallaher IS, et al. Disseminated tumor cells in prostate cancer patients after radical prostatectomy and without evidence of disease predicts biochemical recurrence. *Clin Cancer Res* 2009;15:677–83.
- Klein CA. The systemic progression of human cancer: a focus on the individual disseminated cancer cell—the unit of selection. *Adv Cancer Res* 2003;89:35–67.
- Klein CA, Seidl S, Petat-Dutter K, Offner S, Geigl JB, Schmidt-Kittler O, et al. Combined transcriptome and genome analysis of single micrometastatic cells. *Nat Biotechnol* 2002;20:387–92.
- Choesmel V, Anract P, Holfodt H, Thierry JP, Blin N. A relevant immunomagnetic assay to detect and characterize epithelial cell adhesion molecule-positive cells in bone marrow from patients with breast carcinoma: immunomagnetic purification of micrometastases. *Cancer* 2004;101:693–703.
- Fehm T, Braun S, Muller V, Janni W, Gebauer G, Marth C, et al. A concept for the standardized detection of disseminated tumor cells in bone marrow from patients with primary breast cancer and its clinical implementation. *Cancer* 2006;107:885–92.
- Klein CA, Schmidt-Kittler O, Schardt JA, Pantel K, Speicher MR, Riethmuller G. Comparative genomic hybridization, loss of heterozygosity, and DNA sequence analysis of single cells. *Proc Natl Acad Sci U S A* 1999;96:4494–9.
- Czyz ZT, Hoffmann M, Schlimok G, Polzer B, Klein CA. Reliable Single Cell Array CGH for Clinical Samples. *PLoS ONE* 2014;9:e85907.
- Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* 1998;95:14863–8.
- Schardt JA, Meyer M, Hartmann CH, Schubert F, Schmidt-Kittler O, Fuhrmann C, et al. Genomic analysis of single cytokeratin-positive cells from bone marrow reveals early mutational events in breast cancer. *Cancer Cell* 2005;8:227–39.
- Thierry JP, Acloque H, Huang RY, Nieto MA. Epithelial-mesenchymal transitions in development and disease. *Cell* 2009;139:871–90.
- Watson MA, Ylagan LR, Trinkaus KM, Gillanders WE, Naughton MJ, Weilbaecher KN, et al. Isolation and molecular profiling of bone marrow micrometastases identifies TWIST1 as a marker of early tumor relapse in breast cancer patients. *Clin Cancer Res* 2007;13:5001–9.
- Rao CG, Chianese D, Doyle GV, Miller MC, Russell T, Sanders RA Jr, et al. Expression of epithelial cell adhesion molecule in carcinoma cells present in blood and primary and metastatic tumors. *Int J Oncol* 2005;27:49–57.
- Lazar DC, Cho EH, Luttgen MS, Metzner TJ, Uson ML, Torrey M, et al. Cytometric comparisons between circulating tumor cells from prostate cancer patients and the prostate-tumor-derived LNCaP cell line. *Phys Biol* 2012;9:016002.
- Munz M, Fellinger K, Hofmann T, Schmitt B, Gires O. Glycosylation is crucial for stability of tumour and cancer stem cell antigen EpCAM. *Front Biosci* 2008;13:5195–201.
- Yang Y, Hao J, Liu X, Dalkin B, Nagle RB. Differential expression of cytokeratin mRNA and protein in normal prostate, prostatic intraepithelial neoplasia, and invasive carcinoma. *Am J Pathol* 1997;150:693–704.
- van Leenders GJ, Schalken JA. Epithelial cell differentiation in the human prostate epithelium: implications for the pathogenesis and therapy of prostate cancer. *Crit Rev Oncol Hematol* 2003;46:S3–10.
- Collins AT, Berry PA, Hyde C, Stower MJ, Maitland NJ. Prospective identification of tumorigenic prostate cancer stem cells. *Cancer Res* 2005;65:10946–51.
- Schmelz M, Moll R, Hesse U, Prasad AR, Gandolfi JA, Hasan SR, et al. Identification of a stem cell candidate in the normal human prostate gland. *Eur J Cell Biol* 2005;84:341–54.
- Boon T, Cerottini JC, Van den Eynde B, van der Bruggen P, Van Pel A. Tumor antigens recognized by T lymphocytes. *Annu Rev Immunol* 1994;12:337–65.
- Mecklenburg I, Weckermann D, Zippelius A, Schoberth A, Petersen S, Prang N, et al. A multimer real-time RT-PCR for MAGE-A gene expression allows sensitive detection and quantification of the minimal systemic tumor load in patients with localized cancer. *J Immunol Methods* 2007;323:180–93.
- Mueller P, Carroll P, Bowers E, Moore D II, Cher M, Presti J, et al. Low frequency epithelial cells in bone marrow aspirates from prostate carcinoma patients are cytogenetically aberrant. *Cancer* 1998;83:538–46.
- Riesenberg R, Oberneder R, Kriegmair M, Epp M, Bitzer U, Hofstetter A, et al. Immunocytochemical double staining of cytokeratin and prostate specific antigen in individual prostatic tumour cells. *Histochemistry* 1993;99:61–6.
- Ellis WJ, Pfitzenmaier J, Colli J, Arfman E, Lange PH, Vessella RL. Detection and isolation of prostate cancer cells from peripheral blood and bone marrow. *Urology* 2003;61:277–81.
- Pfitzenmaier J, Ellis WJ, Hawley S, Arfman EW, Klein JR, Lange PH, et al. The detection and isolation of viable prostate-specific antigen positive epithelial cells by enrichment: a comparison to standard prostate-specific antigen reverse transcriptase polymerase chain reaction and its clinical relevance in prostate cancer. *Urol Oncol* 2007;25:214–20.
- Buhring HJ, Muller T, Herbst R, Cole S, Rappold I, Schuller W, et al. The adhesion molecule E-cadherin and a surface antigen recognized by the antibody 9C4 are selectively expressed on erythroid cells of defined maturational stages. *Leukemia* 1996;10:106–16.
- Lammers R, Giesert C, Grunebach F, Marxer A, Vogel W, Buhring HJ. Monoclonal antibody 9C4 recognizes epithelial cellular adhesion molecule, a cell surface antigen expressed in early steps of erythropoiesis. *Exp Hematol* 2002;30:537–45.
- Israels LG, Israels ID. Erythropoiesis: an overview. In: Molineux G, Foote MA, Elliott SG, editors. *Erythropoietins and erythropoiesis*. Basel, Switzerland: Birkhäuser Verlag; 2003. p. 3–14.

35. Fukuda M, Fukuda MN, Papayannopoulou T, Hakomori S. Membrane differentiation in human erythroid cells: unique profiles of cell surface glycoproteins expressed in erythroblasts *in vitro* from three ontogenic stages. *Proc Natl Acad Sci U S A* 1980;77: 3474–8.
36. Villeval JL, Le Van Kim C, Bettaieb A, Debili N, Colin Y, el Maliki B, et al. Early expression of glycophorin C during normal and leukemic human erythroid differentiation. *Cancer Res* 1989;49:2626–32.
37. Attard G, Swennenhuis JF, Olmos D, Reid AH, Vickers E, A'Hern R, et al. Characterization of ERG, AR and PTEN gene status in circulating tumor cells from patients with castration-resistant prostate cancer. *Cancer Res* 2009;69:2912–8.
38. Yu M, Ting DT, Stott SL, Wittner BS, Oszolak F, Paul S, et al. RNA sequencing of pancreatic circulating tumour cells implicates WNT signalling in metastasis. *Nature* 2012;487:510–3.
39. Ozkumur E, Shah AM, Ciciliano JC, Emmink BL, Miyamoto DT, Brachtel E, et al. Inertial focusing for tumor antigen-dependent and -independent sorting of rare circulating tumor cells. *Sci Transl Med* 2013;5:179ra47.
40. Capulli M, Angelucci A, Driouch K, Garcia T, Clement-Lacroix P, Martella F, et al. Increased expression of a set of genes enriched in oxygen binding function discloses a predisposition of breast cancer bone metastases to generate metastasis spread in multiple organs. *J Bone Miner Res* 2012;27:2387–98.
41. Gorr TA, Wichmann D, Pilarsky C, Theurillat JP, Fabrizio A, Laufs T, et al. Old proteins - new locations: myoglobin, hemoglobin, neuroglobin and cytoglobin in solid tumours and cancer cells. *Acta Physiol* 2011;202:563–81.
42. Zhang S, Mercado-Urbe I, Liu J. Generation of erythroid cells from fibroblasts and cancer cells *in vitro* and *in vivo*. *Cancer Lett* 2013; 333:205–12.
43. Shiozawa Y, Pedersen EA, Havens AM, Jung Y, Mishra A, Joseph J, et al. Human prostate cancer metastases target the hematopoietic stem cell niche to establish footholds in mouse bone marrow. *J Clin Invest* 2011;121:1298–312.
44. Ono M, Kosaka N, Tominaga N, Yoshioka Y, Takeshita F, Takahashi RU, et al. Exosomes from bone marrow mesenchymal stem cells contain a microRNA that promotes dormancy in metastatic breast cancer cells. *Sci Signal* 2014;7:ra63.
45. Emara M, Turner AR, Allalunis-Turner J. Hypoxia differentially upregulates the expression of embryonic, fetal and adult hemoglobin in human glioblastoma cells. *Int J Oncol* 2014;44:950–8.
46. Liu W, Baker SS, Baker RD, Nowak NJ, Zhu L. Upregulation of hemoglobin expression by oxidative stress in hepatocytes and its implication in nonalcoholic steatohepatitis. *PLoS ONE* 2011;6:e24363.
47. Heidenreich A, Bastian PJ, Bellmunt J, Bolla M, Joniau S, Mason MD, et al. Guidelines on prostate cancer: European Association of Urology; Arnhem, the Netherlands, 2012.

## **Supplementary information for**

### **Combined genome and transcriptome analysis of single disseminated cancer cells from bone marrow of prostate cancer patients reveals unexpected transcriptomes**

## **Supplementary Methods**

### ***Processing of BM samples***

Upon arrival, the BM sample was washed with Hank's salt solution (HSS; Biochrom AG) to remove fat and thrombocytes. The sample was centrifuged at 170 x *g*, for 10 min at 4°C and supernatant was discarded. The washing was done twice. Next, the cell pellet was resuspended in 9 ml of HSS and overlaid on 6 ml of 65% Percoll solution ( $\rho = 1.083 \text{ g/cm}^3$ ; GE Healthcare). The sample was centrifuged at 1000 x *g*, for 20 min at 4°C, to remove erythrocytes and granulocytes. After centrifugation, the interphase containing mononuclear cells (MNCs) was carefully collected and washed with phosphate buffered saline (PBS). Cell suspension was centrifuged at 500 x *g*, for 10 min at 4°C. The number of MNCs and erythrocytes was determined using a haemocytometer. To enrich the DCC-containing fraction the sample was depleted of the majority of hematopoietic cells using negative immunomagnetic selection. To achieve this, the cell pellet was resuspended in MACS buffer (0.5% BSA, 2 mM EDTA in PBS; 90  $\mu\text{l}/10^7$  MNC). Then, the cell suspension was incubated with APC-conjugated antibodies against CD11b, expressed by monocytes, granulocytes, macrophages, and natural killer cells (10  $\mu\text{l}/10^7$  MNC; antibody concentration was not supplied by manufacturer, Miltenyi Biotec), CD33, expressed by myeloid cells (5  $\mu\text{l}/10^7$  MNC; Invitrogen), and CD45, common antigen of leukocytes (5  $\mu\text{l}/10^7$  MNC; Invitrogen). After 15 min incubation at 4°C, the suspension was washed in MACS buffer. The cell pellet was again re-suspended in MACS buffer (60  $\mu\text{l}/10^7$  erythrocytes). Then,

anti-APC beads (20  $\mu$ l/ $10^7$  MNC; Miltenyi Biotec) and anti-235a beads (glycophorin A, expressed on mature erythroid cells; 20  $\mu$ l/ $10^7$  erythrocytes; Miltenyi Biotec) were added. After 15 min incubation at 4°C, the suspension was washed in MACS buffer. The cell pellet was re-suspended in 2 ml of MACS buffer, run through the 40  $\mu$ m cell sieve, and then run on LS MACS column (Miltenyi Biotec), previously equilibrated with MACS buffer. The column was washed with 9 ml of MACS buffer. Eleven ml of the eluate containing the marker-negative cell fraction were collected on ice and cell number determined on haemocytometer.

### ***Staining of BM samples and cell lines***

On average, two million BM cells were re-suspended in 100  $\mu$ l of blocking solution (10% human AB serum (Biotest) in PBS) and incubated for 15 min under constant rotation at 4°C. Next, 2  $\mu$ l of anti-EpCAM-PE (HEA125; antibody concentration was not supplied by manufacturer, Miltenyi Biotec) antibody were added and incubated for 15 min on the roller at 4°C. After incubation, the antibody solution was removed and the cell pellet washed with PBS. Staining with isotype antibodies (Dako) was performed in 12 cases (10% of all BM samples) and in no case unspecific staining was observed. Remaining samples were not stained with the isotype antibody due to the lack of sufficient number of cells (the same BM samples were used for other *in vivo* and *in vitro* studies). All experiments with cell lines involved included controls using isotype antibodies (Supplementary Figure S2).

Cell lines (WPE-Stem, DU145, PC3, WPE-Int, LNCaP and VCaP) were grown according to instructions from ATCC. At 80% confluence, the medium was discarded, cell monolayer washed with PBS, and cells fixed using 4% formaldehyde. After fixation, cells were washed with PBS. Further immunostaining steps were performed as described above. Final washing step was done in presence of 0,5 ng/ $\mu$ l dilution of DAPI (Roche) in



PBS. The DAPI solution was incubated for 5 min. This solution was replaced with fresh PBS and the cells observed under microscope.

### ***Screening of BM and isolation of single cells and cell pools***

Each BM sample was manually screened for the presence of EpCAM<sup>+</sup> cells on an inverted fluorescent microscope (Olympus or Zeiss), equipped with micromanipulator (Patchman NP2, Eppendorf) and pump (CellTram, Eppendorf). Cell suspension containing  $\sim 0.3 \times 10^6$  of stained BM cells in 30  $\mu$ l was transferred into the chamber of 8-chamber microscope slide (Nunc) containing 170  $\mu$ l of PBS. Single cells with integer morphology and positively stained for EpCAM were isolated using a glass capillary attached to the micromanipulator. After extraction of the single cell, it was transferred into the chamber containing 200  $\mu$ l of PBS, to ensure that only one cell was isolated. Then, selected single cells were isolated manually using a micropipette, by transferring the single cell in 1  $\mu$ l of PBS into microtubes 0.2 ml tubes containing 10 ng tRNA (Roche), dissolved in 4  $\mu$ l of lysis buffer (mTRAP kit, Active Motif). The microtubes were immediately stored at -80°C. tRNA was added to prevent unspecific binding of lysed single cell RNA to the tube wall. In addition, after screening and isolation of single cells, a pool of ~2000-3000 cells was isolated by taking 1  $\mu$ l of cell suspension and transferring it to the microtubes with tRNA and lysis buffer. After isolation of single cells, 1  $\mu$ l of cell-free PBS, from which individual cells were isolated, was also isolated for subsequent WTA as control.

### ***Whole transcriptome amplification (WTA) of single cells***

Single cells in 1  $\mu$ l of PBS were isolated using micromanipulation and deposited in tubes containing 6.4  $\mu$ l of lysis buffer (Active Motif) containing 10 ng of tRNA (Roche), 1  $\mu$ g of protease (Active Motif), and 1  $\mu$ l of 37.5  $\mu$ M solution of biotinylated oligo-dT peptide nucleic acids (PNAs; Active Motif). Proteolytic digestion was performed by incubating the

samples for 10 min at 45°C, followed by inactivation of protease at 75°C for 1 min, and annealing of PNA to poly-A tails of mRNAs, at 22°C for 15 min. PNA-mRNA complexes were precipitated in magnetic force field using streptavidin-conjugated metal beads. While precipitated in magnetic racks, bead pellets were washed with 10 µl of wash buffer 1 (50 mM Tris-HCl, 75 mM KCl, 10 mM DTT, 0.25% Igepal), 20 µl of wash buffer 2 (50 mM Tris-HCl, 75 mM KCl, 10 mM DTT, 0.5% Tween-20), and again with 20 µl of wash buffer 1. These DNA-containing supernatants were transferred to a tube, containing 0.8 µl of polymerized 0.25% polyacrylamide as a carrier, for subsequent precipitation and WGA. Reverse transcription on solid phase was performed for 45 min under rotation at 44°C, in 20 µl reaction mixture containing 0.5 mM of each dNTP (GE Healthcare), 200 U of SuperScript II reverse transcriptase (Invitrogen), 0.25% Igepal, 5 mM DTT, 30 µM of C<sub>15</sub>GTCTAGAN<sub>8</sub> primer, 15 µM of C<sub>15</sub>GTCTAGACTTGAGT<sub>24</sub>VN primer (Metabion), and 1x first strand buffer (Invitrogen). Primers were annealed at room temperature for 10 min, prior to addition of the enzyme. Following reverse transcription, beads were precipitated in magnetic racks and washed in 20 µl of wash buffer 3 (50 mM KH<sub>2</sub>PO<sub>4</sub>, 1 mM DTT, 0.25% Igepal), and resuspended in 10 µl of buffer for tailing (4 mM MgCl<sub>2</sub>, 0.1 mM DTT, 0.2 mM dGTP, 10 mM KH<sub>2</sub>PO<sub>4</sub>). Reaction mixture was overlaid with 40 µl of mineral oil, and the cDNA single strands released from beads by incubating the mixture at 95°C for 5 min, followed by incubation on ice for 3 min. Addition of dGTPs on 5' termini of single stranded cDNAs was performed by adding 10 U of terminal dNTP transferase (TdT; USB-Affymetrix) and incubating the mixture for 60 min at 37°C. After inactivation of TdT at 70°C for 5 min, we added 35 µl of WTA reaction mix 1 (4 µl of buffer I (Expand Long Template, Roche), 3% deionized formamide). Hotstart PCR was performed by incubating the sample to 78°C and adding 5.5 µl of WTA reaction mix 2 (3.2 mM each dNTP, 12 mM TCAGAATTCATGC<sub>15</sub> primer, and 7.5 U of PolMix (Expand Long Template, Roche)). WTA consisted of 40 cycles in MJ Research PCR cycler – 20 cycles of 15 s at 94°C, 30 s at

65°C, and 2 min at 68°C, and 20 cycles with an increase of elongation step for 10 s/cycle, followed by final cycle with 7 min of elongation.

### ***Whole genome amplification (WGA)***

Sixty microliters of DNA- and carrier PAA-containing supernatant were mixed with 120 µl of ice-cold absolute ethanol and left at -20°C to precipitate. Subsequently, the tubes were centrifuged at 4°C for 45 min at 20800 x g. The supernatant was removed, and the pellet was washed with 180 µl of 70% ice-cold ethanol, for 10 min at 18°C in thermomixer (Eppendorf) set on 350 rpm. Next, the tubes were centrifuged at room temperature for 10 min at 20800 x g. These washing steps were repeated twice. After final centrifugation, the pellet was air-dried, and resuspended in 3.5 µl of water, and incubated for 18 h at 18°C in thermomixer (Eppendorf) set on 350 rpm. Subsequent WGA procedure was performed using the Ampli1<sup>TM</sup> WGA Kit (Silicon Biosystems), which is based on the method described previously (10, 13).

### ***Establishing conditions for specific end-point PCR***

Primers were selected based on the cDNA sequence of each gene of interest (48, 49), using Primer3 online tool (50). Specificity of each primer was checked using BLAST (51). This analysis showed that each primer pair detects all isoforms of the corresponding transcript. The only exception was the transcript of *MAGEA2*, where primers detect 6 out of 7 isoforms. Amplification of *MAGEA* genes proved to be very difficult, so we used already published primers (27, 52).

WTA products of different single cell line cells (VCaP, PC3, A431 and BT20) and BM cells were used to establish conditions for PCR amplification of selected transcripts (Supplementary Figure S11). These WTA products were from our collection of cell line WTA products. All annealing temperatures were selected based on results of gradient



PCR. PCR was performed in 10 µl of water solution containing 0.1 µl of WTA cDNA, 10 mM Tris-HCl, 50 mM KCl, 1 mM MgCl<sub>2</sub>, 0.1 mM equimolar mixture of dNTPs (GE Healthcare), 0.4 pM of each primer (Supplementary Table S1), 0.5 µg/µl of BSA (Roche), and 0.5 U of Taq DNA polymerase (Pan Biotech). Thermal cycling was performed as follows: 1<sup>st</sup> phase: denaturation at 94°C for 2 min, annealing (transcript-specific, Supplementary Table S1) for 30 sec, and extension at 72°C for 2 min for the first cycle; 2<sup>nd</sup> phase: denaturation at 94°C for 15 sec, annealing for 30 sec, and extension at 72°C for 20 sec for 14 cycles; 3<sup>rd</sup> phase: denaturation at 94°C for 15 sec, annealing for 30 sec, and extension at 72°C for 30 sec for 25 cycles; 4<sup>th</sup> phase: extension at 72°C for 2 min. All PCR products were electrophoresed through a 3% agarose gel and stained with ethidium bromide for visualization under ultraviolet light. To confirm the identity of amplified fragments, we digested all PCR products with adequate restriction endonuclease (NEB; Supplementary Figure S11 and Supplementary Table S1).

Restriction digestion was performed in a 30 µl water solution containing 15 µl of PCR reaction product, 3 µl of corresponding restriction endonuclease (Supplementary Table S1) and 3 µl of adequate reaction buffer (NEB; NEBuffer 1 for Tsp509I, NEBuffer 3 for Bsrl and NEBuffer 4 for Hpy188I, Hpy188III, HpyCH4IV, HpyCH4V, HaeIII and NlaIII). For enzymes Hpy188III and NlaIII reaction was supplemented with 10 ng/µl of BSA (Roche). Incubation for Hpy188I, Hpy188III, HpyCH4IV, HpyCH4V, and NlaIII was at 37°C for 3 h followed by inactivation at 65°C for 20 min, for enzyme HaeIII at 37°C for 3 h and inactivation at 80°C for 20 min, for Tsp509I at 65°C for 3 h without inactivation and for Bsrl at 65°C for 3 h and inactivation at 80°C for 20 min.

Additionally, all primers were tested on human genomic DNA (Invitrogen) and failed to produce the correct fragment, demonstrating their specificity for cDNA.

### **Supplementary References**

48. Maglott D, Ostell J, Pruitt KD, Tatusova T. Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res.* 2011;39:D52-7.
49. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. GenBank. *Nucleic Acids Res.* 2010;38:D46-D51.
50. Rozen S, Skaletsky HJ. Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S, editors. *Bioinformatics Methods and Protocols*. Totowa, USA: Humana Press; 2000. p. 365-86.
51. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990;215:403-10.
52. Kufer P, Zippelius A, Lutterbuse R, Mecklenburg I, Enzmann T, Montag A, et al. Heterogeneous expression of MAGE-A genes in occult disseminated tumor cells: a novel multimarker reverse transcription-polymerase chain reaction for diagnosis of micrometastatic disease. *Cancer Res.* 2002;62:251-61.

## Supplementary Tables

### Supplementary Table S1. List of analyzed transcripts.

GeneBank name and common synonym, sequence of primers, annealing temperatures (Ta) for PCR, restriction endonucleases used to identify given fragments and the size of fragments after digestion of PCR product.

Gene	GenBank identifier	Primer	Sequence (5'-3')	Fragment size (bp)	Ta (°C)	Restriction enzyme	Digestion fragments (bp)
<i>EPCAM</i>	NM_002354.2	Forward Reverse	AAAGTTTGC GGACTGCACTT AGCCACATCAGCTATGTCCA	161	55	Hpy188III	94 + 67
<i>KRT8</i> (CK8)	NM_002273.3	Forward Reverse	GCGTACAGAGATGGAGAACGA AGCTCCCGGATCTCCTCTT	150	58	Hpy188III	116 + 34
<i>KRT18</i> (CK18)	NM_000224.2	Forward Reverse	GGAAGATGGCGAGGACTTTA CCTGCTTCTGCTGGCTTAAT	157	55	HpyCH4V	102 + 55
<i>KRT19</i> (CK19)	NM_002276.4	Forward Reverse	AGCCGGACTGAAGAATTGAA TTCTGCCAGTGTGTCTTCCA	162	62	HaeIII	127 + 35
<i>KRT5</i> (CK5)	NM_000424.3	Forward Reverse	GGAGCTCATGAACACCAAGC TGGTCCAACTCCTTCTCCAC	103	64	HaeIII	80 + 23
<i>KRT6a</i> (CK6a)	NM_005554.3	Forward Reverse	GTGAGGAGTGCAGGCTGAAT CCAAGACCACTGCCATAGGA	154	64	HaeIII	112 + 42
<i>KRT14</i> (CK14)	NM_000526.4	Forward Reverse	CGCAGTCATCCAGAGATGTG GGGATCTTCCAGTGGGATCT	182	62	HpyCH4V	120 + 62
<i>KLK3</i> (PSA)	NM_001648.2	Forward Reverse	TTGTCTTCCTCACCTGTCC TTTTGTTCCTGATGCAGTGG	198	58	Hpy188I	141 + 57
<i>PTPRC</i> (CD45)	NM_080923.2	Forward Reverse	TTAGGGACACGGCTGACTTC GCTTTGCCCTGTCAAAATA	106	62	NlaIII	72 + 34
<i>CD34</i>	NM_001773.2	Forward Reverse	AGGCCTCAGTGTCTACTGCTG TAGGATCCCCAGCTTTTCA	105	62	Tsp509I	64 + 41
<i>CD33</i>	NM_001772.3	Forward Reverse	GGCACCAACCTGACCTGT TCTTGTTTCCCTGAGCCATC	137	62	BsrI	105 + 32
<i>CD19</i>	NM_001770.5	Forward Reverse	CCTTGTGGGCATTCTTCATC CACGTTCCCGTACTGGTTCT	136	62	Hpy188III	97 + 39
<i>GYPC</i>	NM_002101.3	Forward Reverse	GAGAAGCCCCAACAGCAC AATCACACCTGCAATGACGA	187/130*	60	Hpy188III	114 + 73
<i>SCL4A1</i> (Band 3)	NM_000342.3	Forward Reverse	TCTTCAGGAACGTGGAGCTT CTGCTTTTCCTTGGAAAGGTG	172	58	NlaIII	98 + 74
<i>HBA2</i>	NM_000517.4	Forward Reverse	GGCTCTGCCCAGGTTAAGG AGGCAGTGGCTTAGGAGCTT	164	60	HpyCH4IV	65 + 99
<i>MAGEA2</i>	NM_175743.1	Forward Reverse	CATTGAAGGAGAAGATCTGCCT CAGGCTTGCACTGCTGACTC	119	58	NlaII	86 + 33
<i>MAGEA4</i>	NM_001011548.1	Forward Reverse	CACCAAGGAGAAGATCTGCCT CAGGCTTGCACTGCTGACTCT	118	58	NlaII	85 + 33
<i>ACTB</i>	NM_001101.3	Forward Reverse	GTGGGGCGCCCCAGGCACCA CTCCTTAATGTACGCACGATTTCC	540	58	n.a.	n.a.
<i>EEF1A1</i>	NM_001402.5	Forward Reverse	TGCCCCAGGACACAGAGACT CTGTGTCGGGGTTGTAGCCA	290	58	n.a.	n.a.
<i>GAPDH</i>	NM_002046.3	Forward Reverse	AATCCCATCACCATCTTCCAG GCCATCACGCCACAGTTTCC	380	58	n.a.	n.a.

\*This primer pair can amplify two *GYPC* isoforms. bp – base pair; n.a. – not applicable.

**Supplementary Table S2.** Detection rate of EpCAM<sup>+</sup> cells in BM samples from M0- and M1-stage PCa patients and controls (N).

The table also lists the number of isolated EpCAM<sup>+</sup> single cells (SC) and clusters of inseparable EpCAM<sup>+</sup> cells, and the number of cells that passed quality control (QC) after WTA and went to downstream transcriptome analysis.  $\Sigma$  – sum.

Screened BM samples	M0		N	M1	
	105		18	2	
EpCAM <sup>+</sup> cells detected in	65 (62%)		10 (56%)	2 (100%)	
Number of isolated EpCAM <sup>+</sup>	SC	clusters	SC	SC	clusters
	222	5	51	12	4
Number of samples selected	119 (54%)	5 (100%)	28 (55%)	7 (58%)	3 (75%)
for transcriptome analysis	$\Sigma$ : 55%		$\Sigma$ : 55%	$\Sigma$ : 62%	
after WTA QC					

**Supplementary Table S3.** Number and frequencies of samples (EpCAM<sup>+</sup> single cells or cell pools), from M0- and M1-stage prostate cancer patients and controls, expressing analyzed transcripts.

Only transcript expression of single cells from M0-stage patients and controls, and of cell pools from M0-stage patients and controls is compared. All differences were not significant, unless indicated otherwise (Pearson Chi-square or Fisher's exact test).

	EpCAM <sup>+</sup> single cells M0 PCa n (%)	EpCAM <sup>+</sup> single cells M1 PCa n (%)	EpCAM <sup>+</sup> single cells control BM n (%)	pools M0 PCa n (%)	pools control BM n (%)
<b>Total nr. of samples</b>	124	10	28	10	10
<b><i>EPCAM</i></b>	109 (88) <sup>a</sup>	8 (80)	14 (50) <sup>a</sup>	8 (80)	8 (80)
<b><i>KRT8</i></b>	37 (30) <sup>b</sup>	5 (50)	1 (4) <sup>b</sup>	10 (100)	10 (100)
<b><i>KRT18</i></b>	33 (27) <sup>c</sup>	2 (20)	2 (7) <sup>c</sup>	9 (90)	9 (90)
<b><i>KRT19</i></b>	6 (5)	1 (10)	0 (0)	2 (20)	0 (0)
<b><i>KRT14</i></b>	1 (1)	0 (0)	0 (0)	0 (0)	0 (0)
<b><i>KRT6a</i></b>	3 (2)	1 (10)	0 (0)	0 (0)	0 (0)
<b><i>KRT5</i></b>	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
<b><i>KLK3</i></b>	9 (7)	1 (10)	0 (0)	1 (10)	0 (0)
<b><i>MAGEA2</i></b>	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
<b><i>MAGEA4</i></b>	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
<b><i>PTPRC</i></b>	23 (19)	2 (20)	6 (21)	10 (100)	6 (60)
<b><i>CD34</i></b>	7 (6)	0 (0)	3 (11)	8 (80)	3 (30)
<b><i>CD33</i></b>	8 (7)	0 (0)	1 (4)	9 (90)	4 (40)
<b><i>CD19</i></b>	5 (4)	0 (0)	2 (7)	8 (80) <sup>d</sup>	2 (20) <sup>d</sup>
<b><i>GYPC</i></b>	65 (52)	2 (20)	14 (50)	10 (100)	8 (80)
<b><i>SCL4A1</i></b>	68 (55)	1 (10)	14 (50)	10 (100)	9 (90)
<b><i>HBA2</i></b>	115 (93)	8 (80)	24 (86)	10 (100)	10 (100)

<sup>a</sup> p<0.001, Pearson Chi-square

<sup>b</sup> p<0.01, Pearson Chi-square

<sup>c</sup> p<0.05, Pearson Chi-square

<sup>d</sup> p<0.05, Fisher's exact test

One pool from a M1-stage patient that was also tested is not included, but can be found in Supplementary Figure S8E

**Supplementary Table S4.** Distribution of EpCAM<sup>+</sup> single cells expressing or co-expressing *EPCAM* and *KRT* transcripts among cells from (a) M0- and (c) M1-stage PCa patients and (b) control patients (N).  $\Sigma$  – sum

(a) M0	<i>KRT</i> <sup>+</sup>	<i>KRT</i> <sup>−</sup>	$\Sigma$
<i>EPCAM</i> <sup>+</sup>	45 (36%) <sup>a</sup>	64 (52%) <sup>b</sup>	109 (88%)
<i>EPCAM</i> <sup>−</sup>	7 (6%)	8 (6%)	15 (12%)
$\Sigma$	52 (42%)	72 (58%)	124

(b) N	<i>KRT</i> <sup>+</sup>	<i>KRT</i> <sup>−</sup>	$\Sigma$
<i>EPCAM</i> <sup>+</sup>	0 <sup>c</sup>	14 (50%) <sup>d</sup>	14 (50%)
<i>EPCAM</i> <sup>−</sup>	2 (7%)	12 (43%)	14 (50%)
$\Sigma$	2 (7%)	26 (93%)	28

(c) M1	<i>KRT</i> <sup>+</sup>	<i>KRT</i> <sup>−</sup>	$\Sigma$
<i>EPCAM</i> <sup>+</sup>	4 (40%)	4 (40%)	8 (80%)
<i>EPCAM</i> <sup>−</sup>	1 (10%)	1 (10%)	2 (20%)
$\Sigma$	5 (50%)	5 (50%)	10

a,b vs. c,d -  $p < 0.002$ , Fisher's exact test

## **Supplementary Figure legends**

**Supplementary Figure S1.** Workflow employed in this study. From each isolated single cell or cluster/pool, total mRNA and genomic DNA were isolated, separated and subjected to global amplification.

**Supplementary Figure S2.** Photographs of the prostate epithelial cell lines shown in Supplementary Figure S4, incubated with isotype antibodies. Cell name is given in the lower right corner of the respective "bright field" images. All photographs are taken with exposure time of 900 ms. Scale bar: 50  $\mu$ m. No unspecific binding was observed.

**Supplementary Figure S3.** Box-Whisker plots containing cell numbers before immunomagnetic selection (IMS input) and after immunomagnetic selection (IMS output), numbers of cells screened for the presence of EpCAM<sup>+</sup> cells and numbers of isolated EpCAM<sup>+</sup> cells/cell clusters per bone marrow sample as indicated on the X-axis. The cell numbers are indicated on the Y-axis in logarithmic scale. Box stretches from 25<sup>th</sup> to 75<sup>th</sup> percentile, with the line in the box representing the median. Whiskers denote minimal and maximal values. Median, mean, minimal and maximal values are shown under the graph.

**Supplementary Figure S4.** Photographs of the prostate epithelial cell lines stained for EpCAM. Cell line name is given in the lower right corner of the respective "bright field" images. All photographs are taken with exposure time of 900 ms. Photographs under "EpCAM (bright. +15%/contr. + 25%)" are manipulated using image processing software in a way that brightness and contrast of the original image were increased as indicated. Scale bar: 50  $\mu$ m.

**Supplementary Figure S5.** Photographs and flowcytometric analysis of the bone marrow sample of a control patient. The sample was split in four parts. Two percent of DU145 cells were spiked into two parts. One spiked and one non-spiked sample were stained with EpCAM-PE antibody, while the remaining two samples were stained with an isotype-PE antibody. There were no stained cells both in the non-spiked (A-D) and spiked (E-H; DU145 cells are indicated with green arrowheads) samples incubated with the isotype-PE antibody. Conversely, weakly EpCAM<sup>+</sup> BM cells (gray arrowheads) could be seen in non-spiked (I-L) and spiked (M-P; DU145 cells are indicated with green arrowheads) samples stained with the EpCAM-PE antibody. Although the EpCAM staining intensity of DU145 cells was slightly higher, only 17% of the spiked-in DU145 population could be identified on basis of their higher fluorescence intensity (Q).

**Supplementary Figure S6.** Frequencies of cell pools expressing analyzed transcripts. The identity of the transcripts is indicated on the X-axis. Results for *KRT14*, *KRT6a*, *KRT5*, *MAGEA2* and *MAGEA4* are not shown, since the transcripts were not detected in any sample. Exact numbers and percentages are given in Supplementary Table 3. M0-stage (grey bars) prostate cancer patients and control patients (white bars). The asterisk indicates the level of significance ( $p < 0.05$ ). n indicates the number of analyzed pools.

**Supplementary Figure S7.** Quality control of WTA and overview of epithelial transcript expression. **A.** Quality control (QC) PCR analysis of WTA products of cells from each sample group - M0- (grey bars) and M1-stage (black bars) prostate cancer patients and control patients (white bars). Labels on the X-axis indicate the number of QC PCR transcripts (*ACTB*, *GAPDH*, *EEF1A1*) for which the samples were positive. The Y-axis indicates the frequency of cells in each category. n indicates the number of analyzed single cells. Absolute numbers are given below the graph. **B.** Frequencies of cells



expressing analyzed transcripts or groups of transcripts within each investigated group - M0- (grey bars) and M1-stage (black bars) prostate cancer patients and control patients (white bars). The identity of transcripts is on the X-axis. “Epithelial marker” represents expression of at least one of the following transcripts: *EPCAM*, *KRT8*, *KRT18*, *KRT19*, *KRT14*, *KRT6a*, and *KRT5*. “KRT” represents expression of at least one of the KRT transcripts (*KRT8*, *KRT18*, *KRT19*, *KRT14*, *KRT6a*, and *KRT5*). Asterisks indicate the level of significance: \*\* –  $p < 0.01$ , \*\*\* –  $p < 0.001$ ; n indicates the number of analyzed single cells. Absolute numbers are given below the graph. **C.** The frequency of cells from M0-stage patients expressing individual KRT transcripts, as indicated on the X-axis. Only cells expressing at least one KRT transcript are included. n indicates the number of analyzed single cells. Absolute numbers are given below the graph.

**Supplementary Figure S8.** Results of PCR analysis of all samples. PCR profiling of selected transcripts in all EpCAM<sup>+</sup> cells/clusters from M0-stage prostate cancer patients (**A**), controls (**B**) and M1-stage prostate cancer patients (**C**), and in cell pools from controls (**D**), M1- (**E**) and M0-stage patients (**F**). Green colour and number “1” indicate that the transcript was detected, while red colour and number “0” indicate that it was not detected. Results for *KRT5*, *MAGEA2* and *MAGEA4* are not shown, since the transcripts were not detected in any sample. Sample identifiers are given on the left side of each panel. Letter ‘c’ indicates that the sample is composed of the cluster of cells. Identifiers of analyzed transcripts are given on top of each panel.

**Supplementary Figure S9.** Analysis of WTA controls for the presence of HBA2 transcript. In total 40 sample were analysed and all were negative. Here, eight samples are shown.

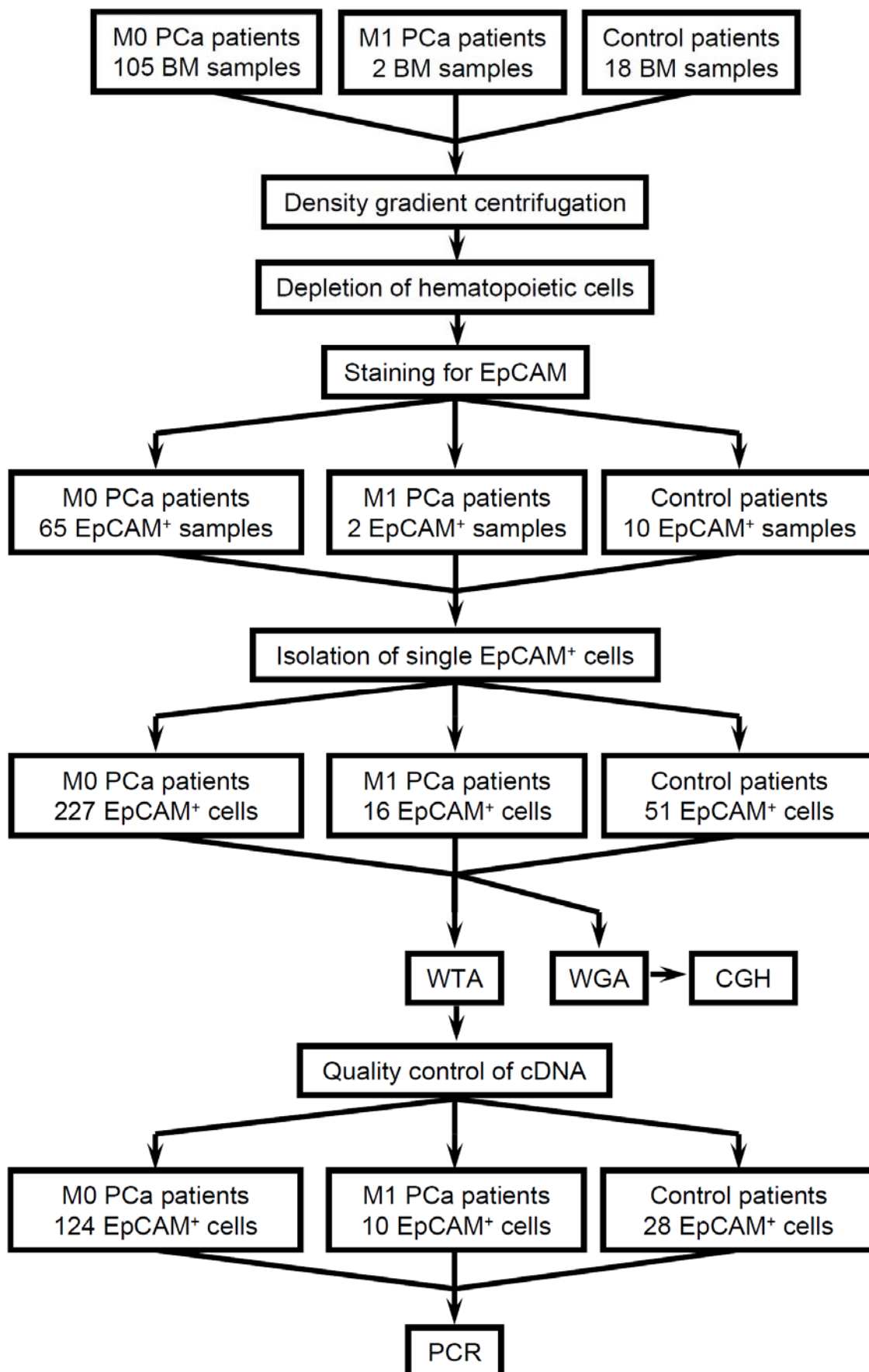
**Supplementary Figure S10.** cCGH and aCGH profiles (as indicated on each individual profile) of all analysed single cells. **A** – Profiles of unamplified genomic DNA of VCaP cell line and of amplified single VCaP cell genome. **B** – Profiles of two individual peripheral blood leukocytes (PBLs) of a healthy donor. **C** – CGH profiles of four EpCAM<sup>+</sup> single cells isolated from BM of cancer-free control patients. These cells contain no aberrations. Results of the transcript profiling of each sample is given under the CGH profile. **D** – CGH profiles of EpCAM<sup>+</sup> single cells without genomic aberrations. The cells were isolated from BM of M0-stage PCa patients. False positive signals sometimes visible in centromeric regions are the consequence of insufficient blocking of the repetitive sequences and are discarded from the analysis. Results of the transcript profiling of each sample is given under the CGH profile (green and 1 indicate that the transcript was detected in the sample, while red and 0 indicate that the transcript was not detected). **E** – CGH profiles of EpCAM<sup>+</sup> single cells/clusters with genomic aberrations. The cells were isolated from BM of M0- and M1-stage PCa patients. These cells display aberrant genomes. Characteristic aberrations detected in genomes of these cells are indicated with arrowheads. Results of the transcript profiling of each sample is given under the CGH profile (green and 1 indicate that the transcript was detected in the sample, while red and 0 indicate that the transcript was not detected). cCGH – CGH on metaphase spreads, aCGH – CGH on DNA microarrays.

**Supplementary Figure S11.** Establishing conditions for specific amplification of selected fragments by PCR and restriction digestion of these fragments. For each transcript the size of PCR amplicon, and digestion fragments is indicated on the left side, while the size of DNA size marker bands is indicated on the right. The sample used for optimizing the conditions is indicated below each panel, and it majority of cases was a single cell (sc) WTA product. Restriction enzyme used for digestion is indicated above the well containing

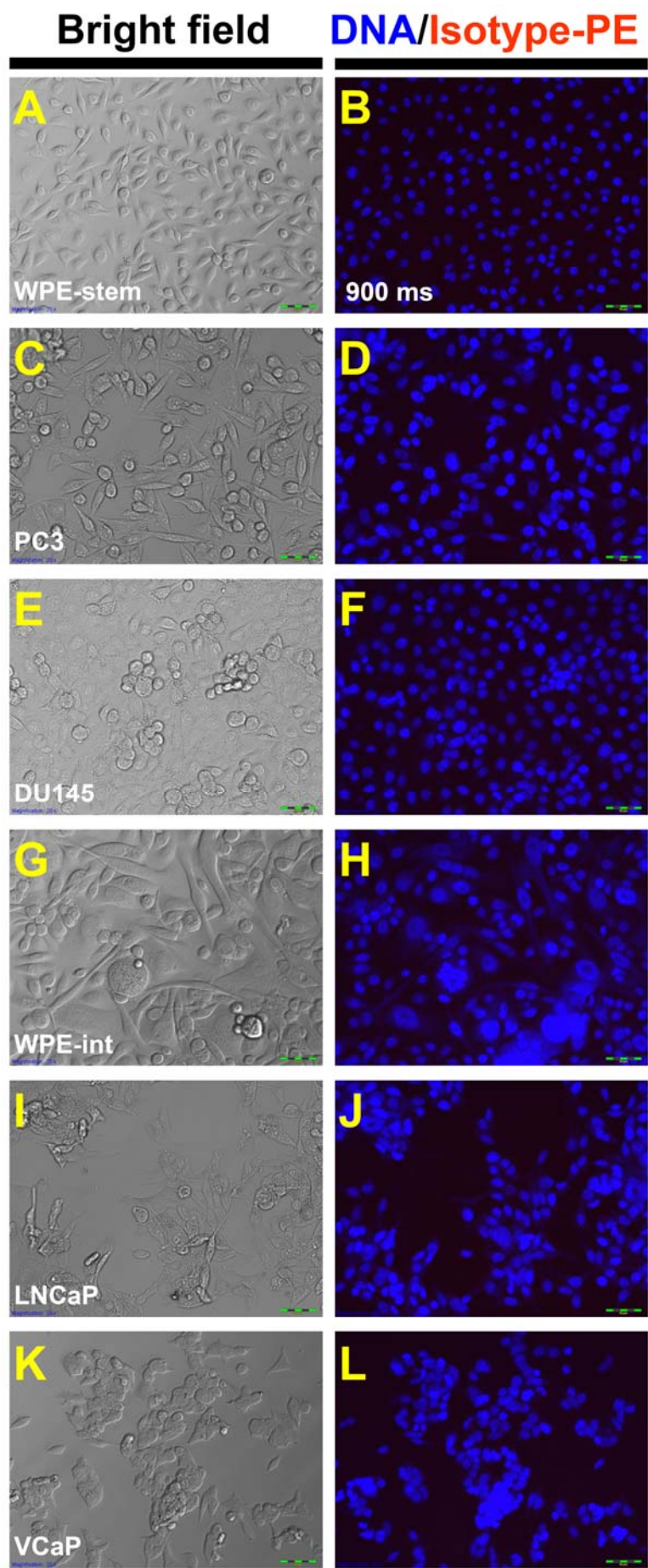
the digested fragments. "No RE" indicates lanes where undigested fragment were run on the gel. "NT" indicates no template control of PCR reaction. The panels, where the lanes between test samples and DNA size marker or NT were removed by cropping the image, are marked by asterisk and separated by a white line.

## Supplementary Figures

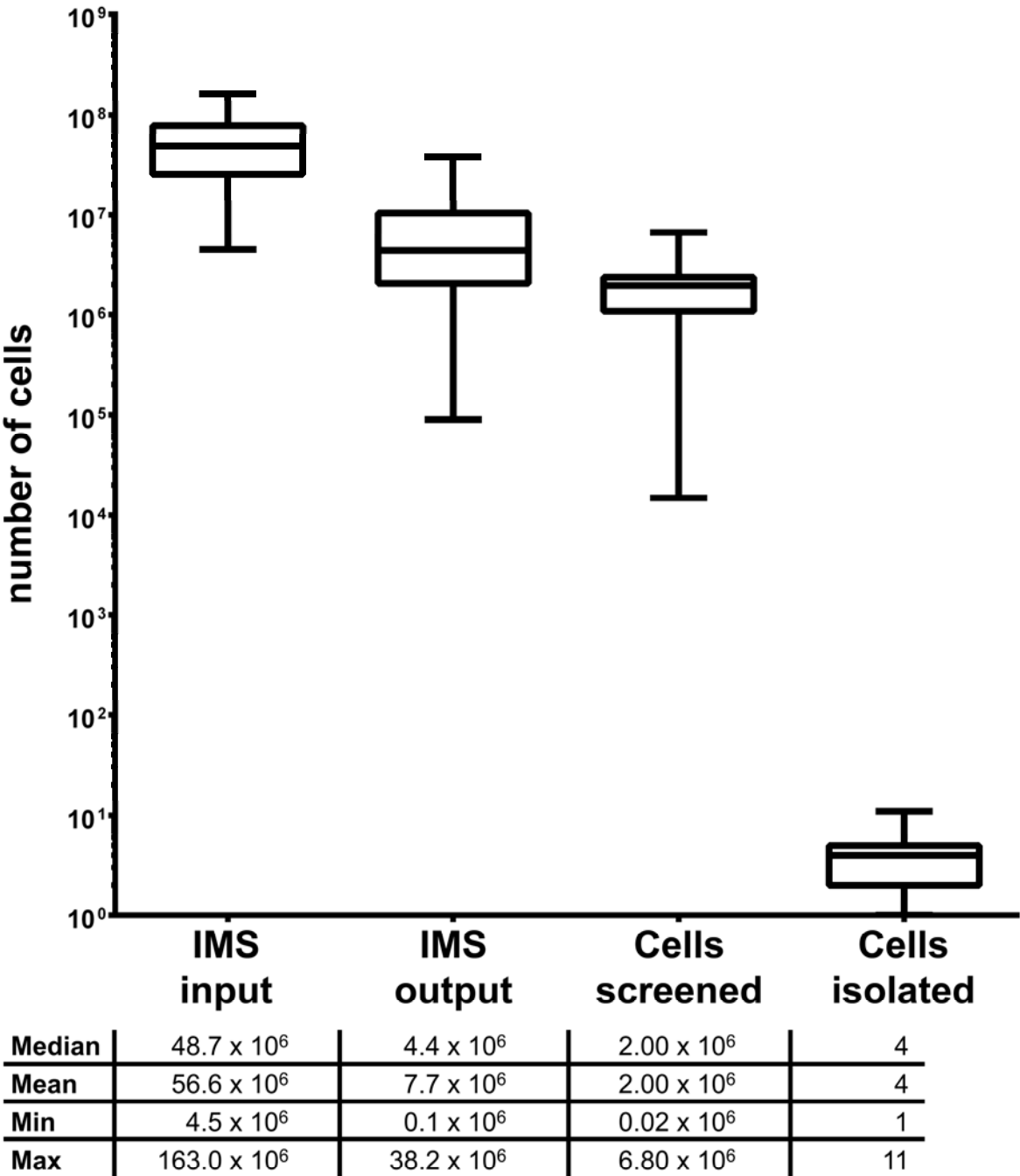
Supplementary Figure S1



Supplementary Figure S2

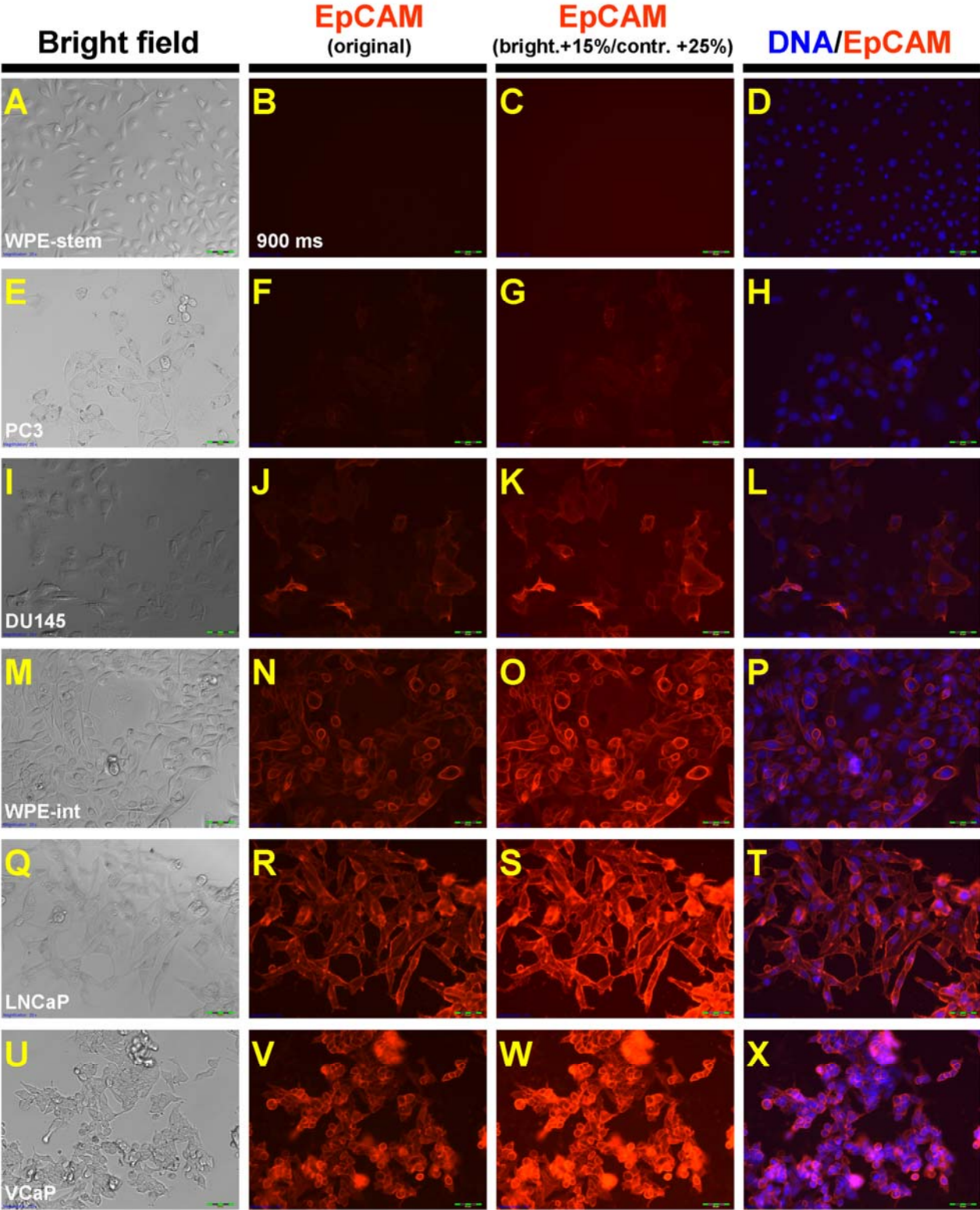


Supplementary Figure S3

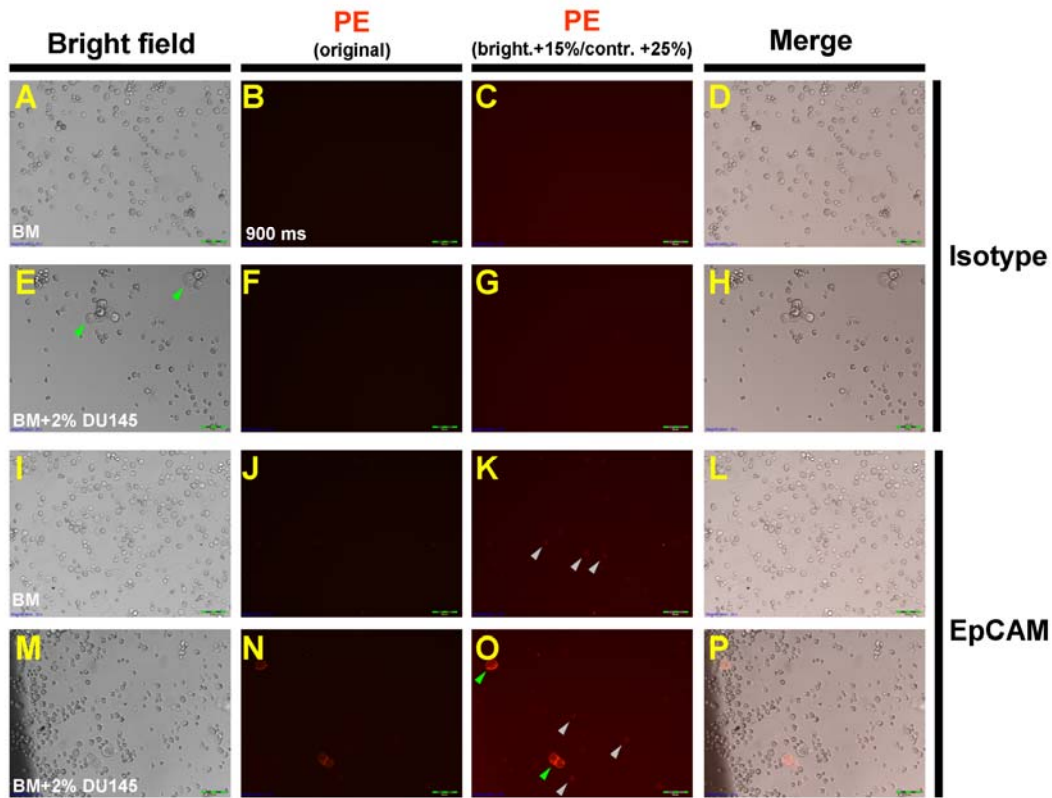




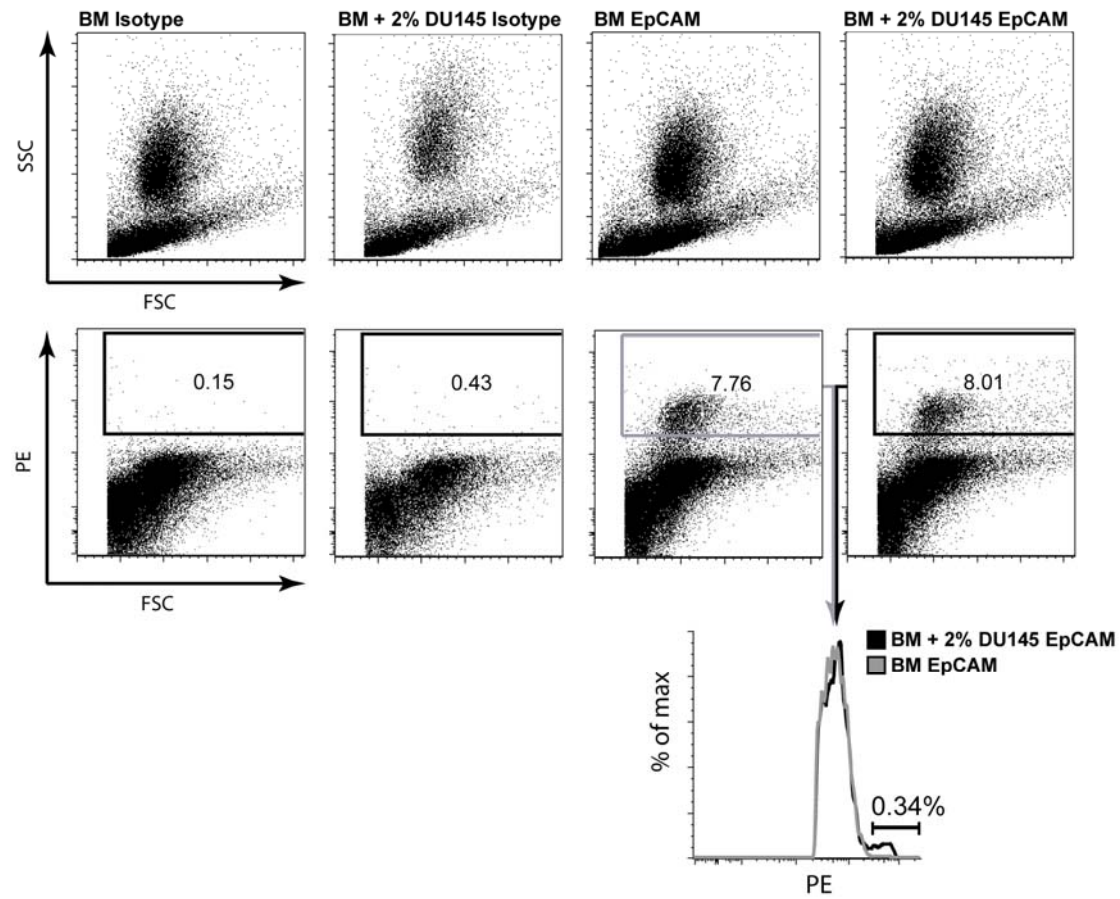
Supplementary Figure S4



Supplementary Figure S5

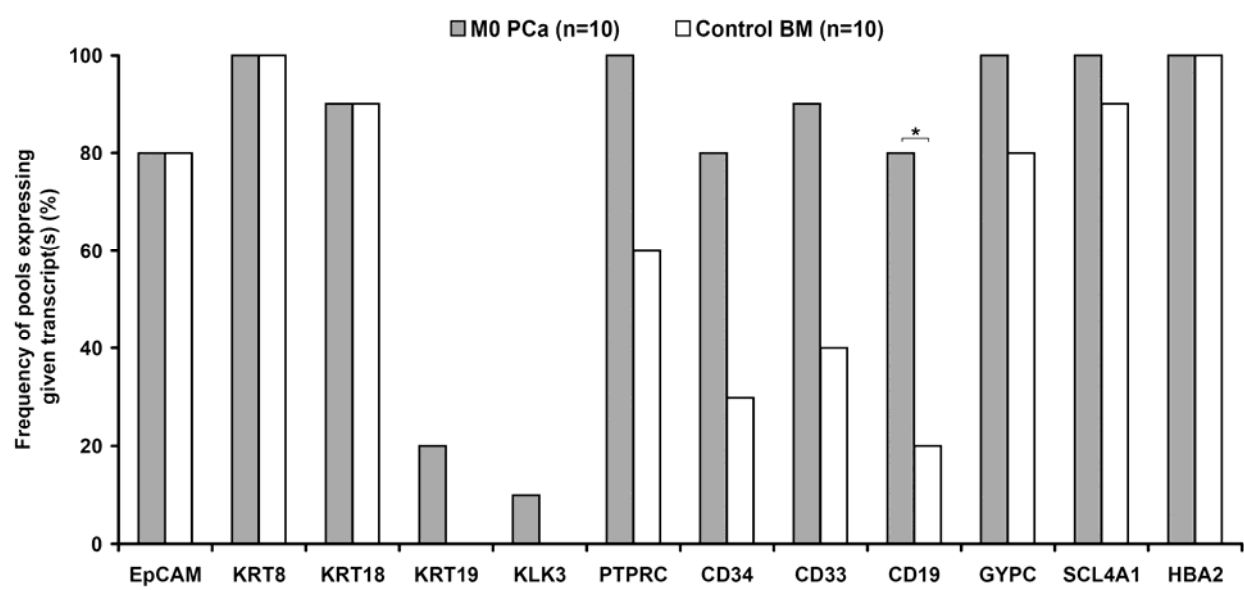


Q

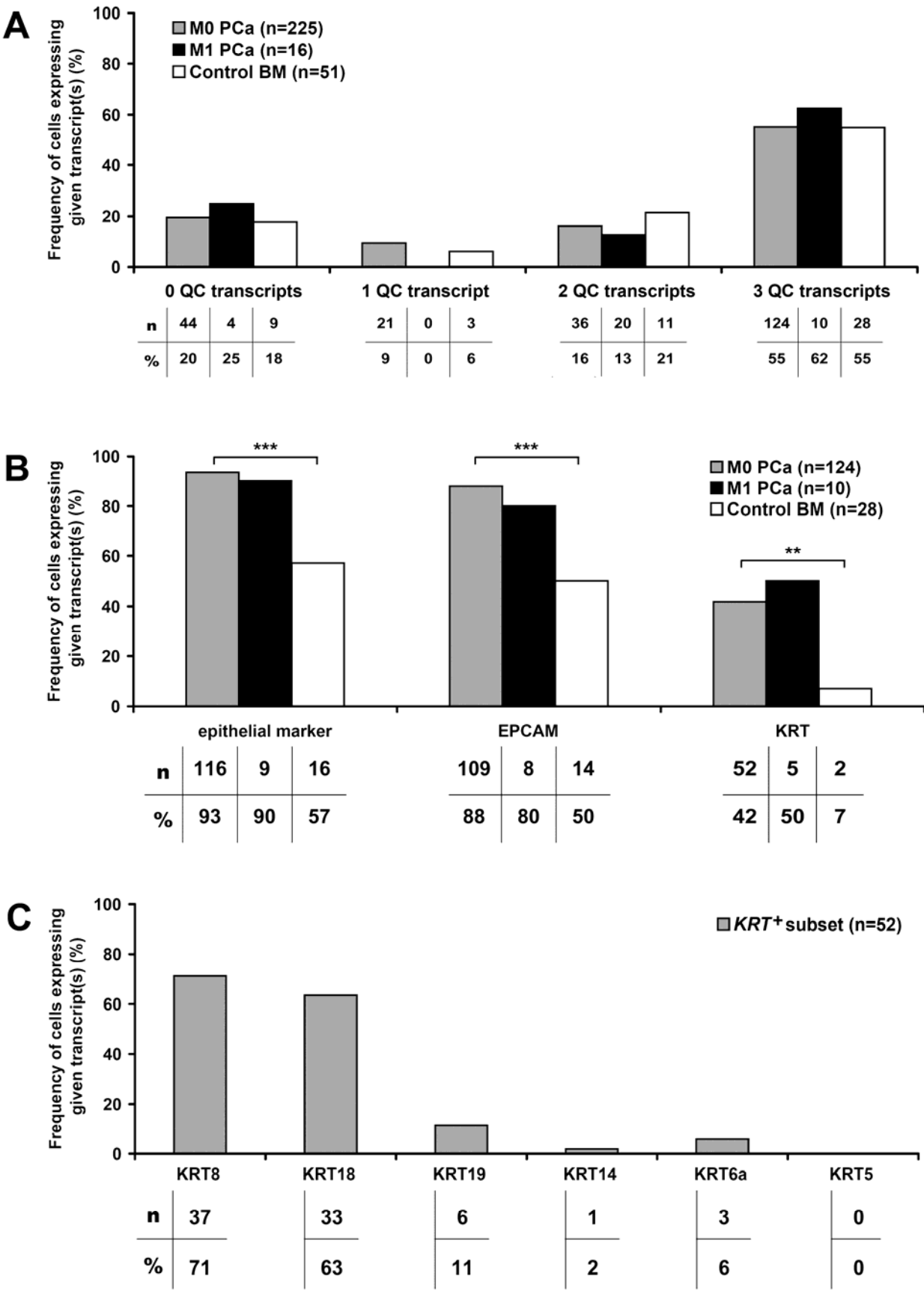




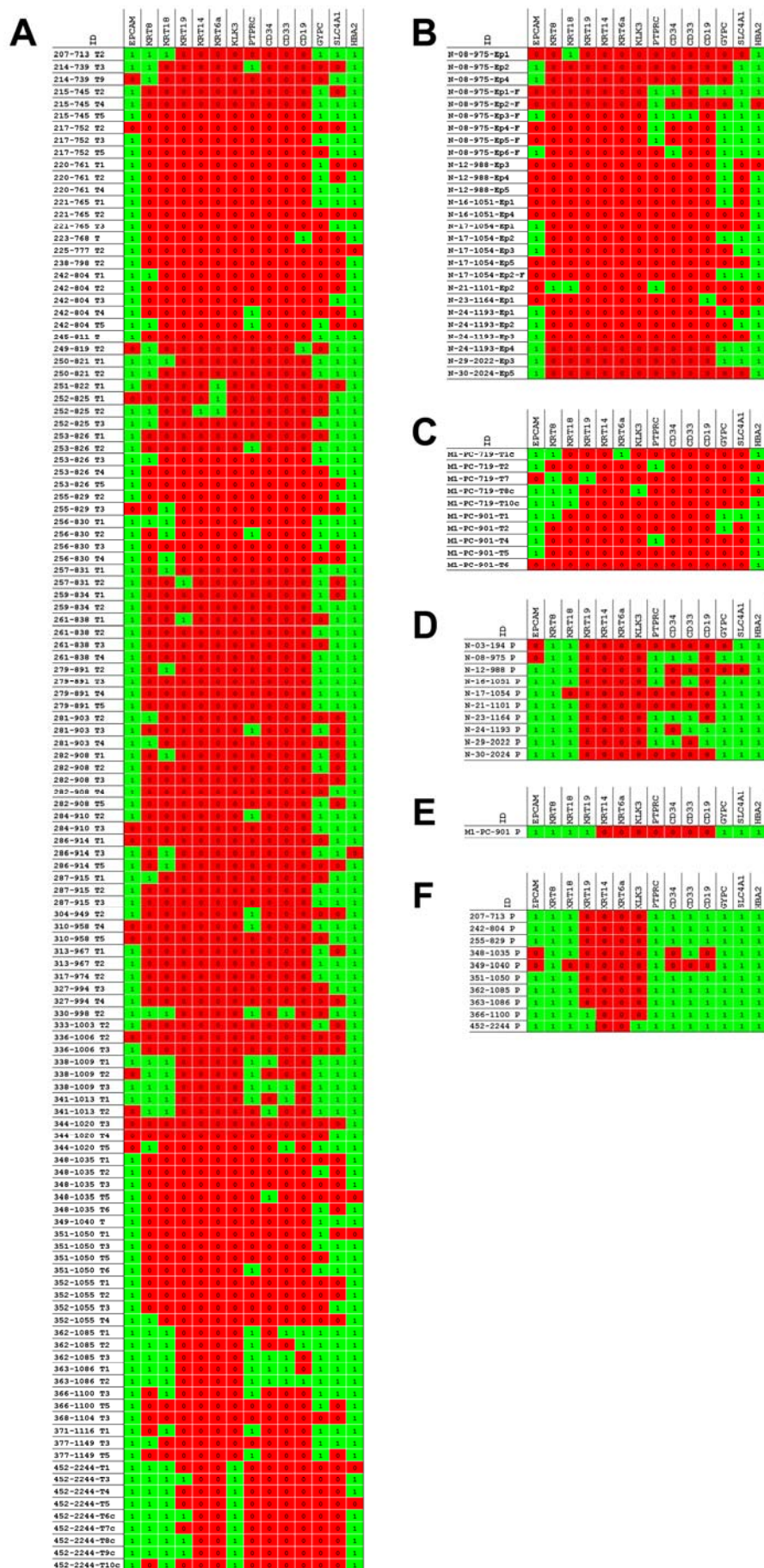
Supplementary Figure S6



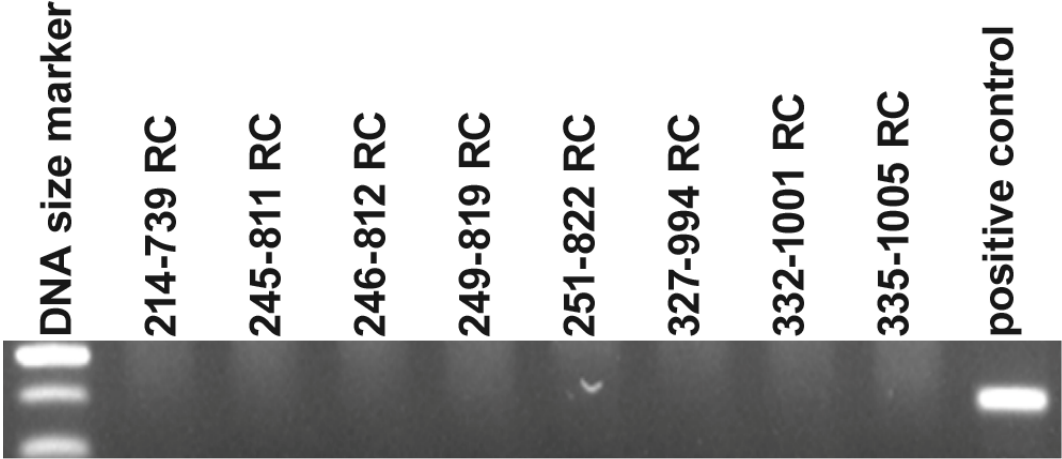
Supplementary Figure S7



### Supplementary Figure S8

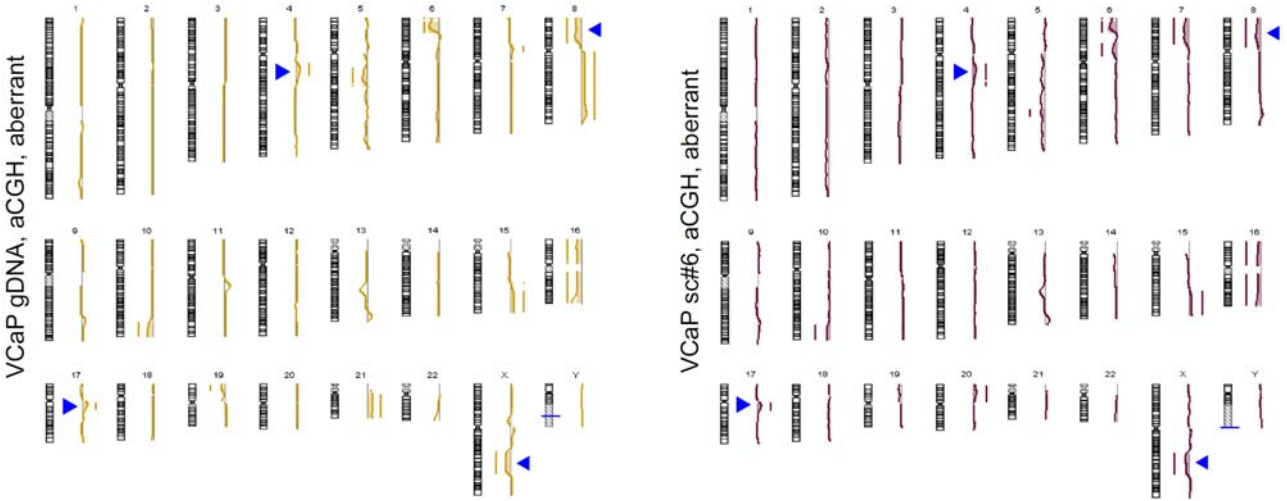


Supplementary Figure S9



Supplementary Figure S10

A

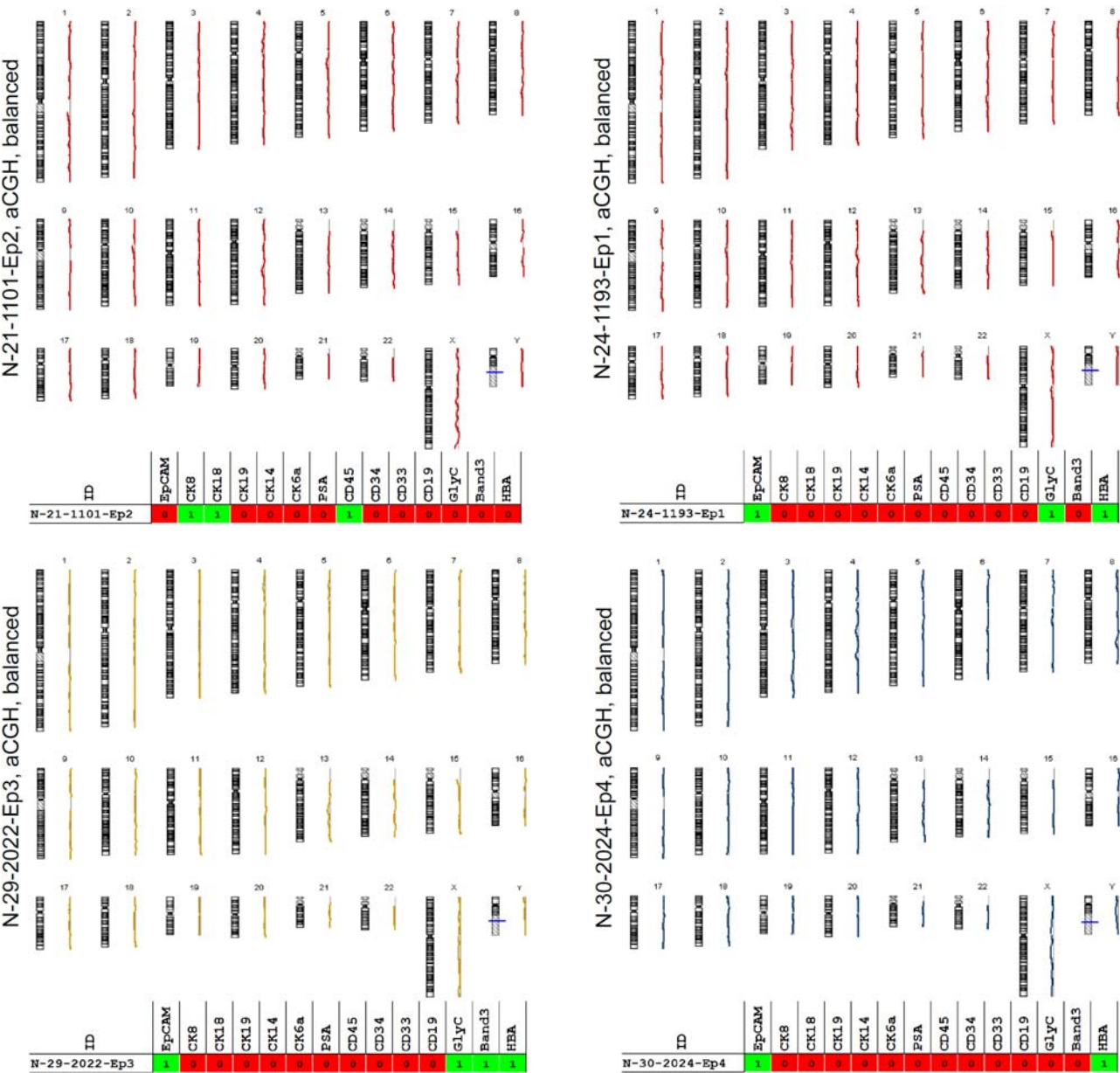


B



Supplementary Figure S10 (continued)

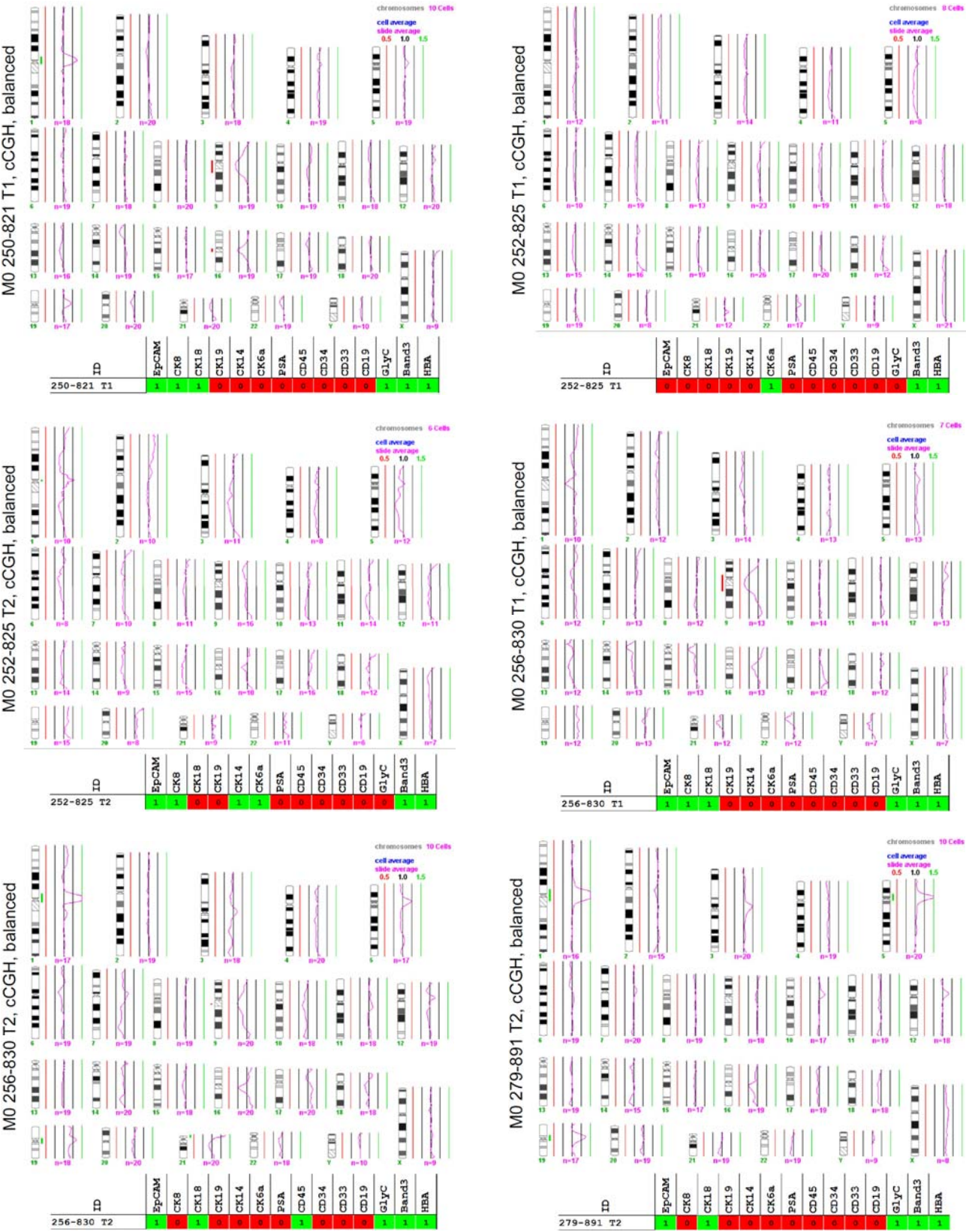
C





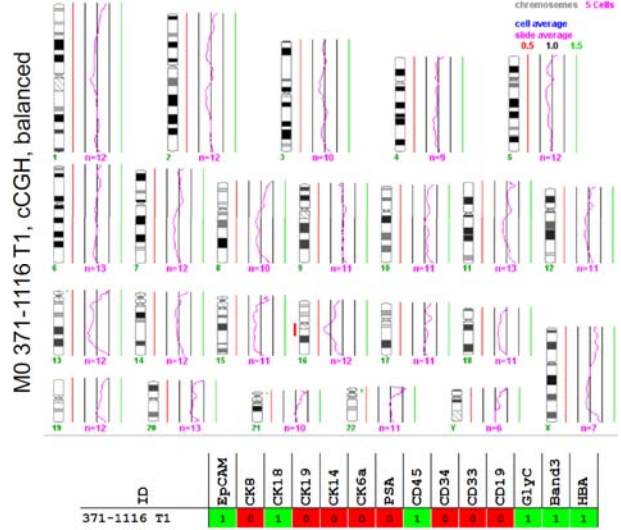
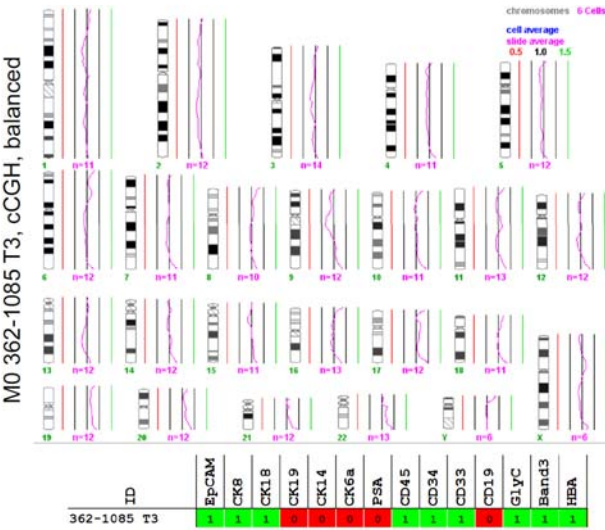
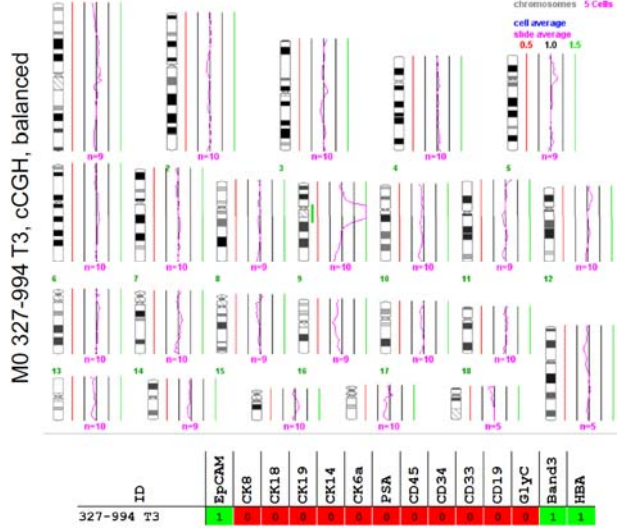
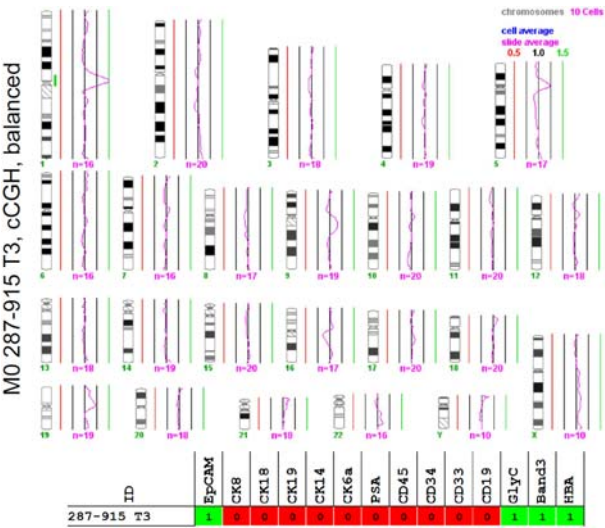
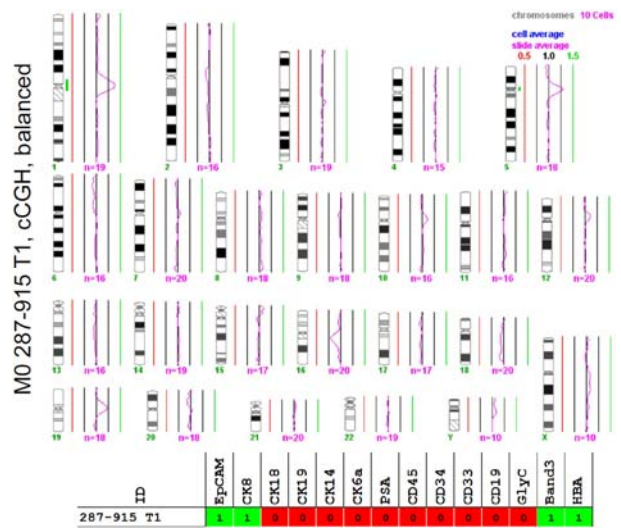
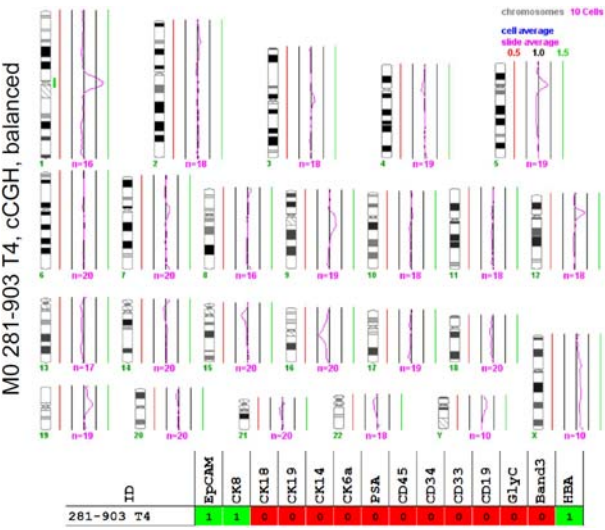
Supplementary Figure S10 (continued)

D



Supplementary Figure S10 (continued)

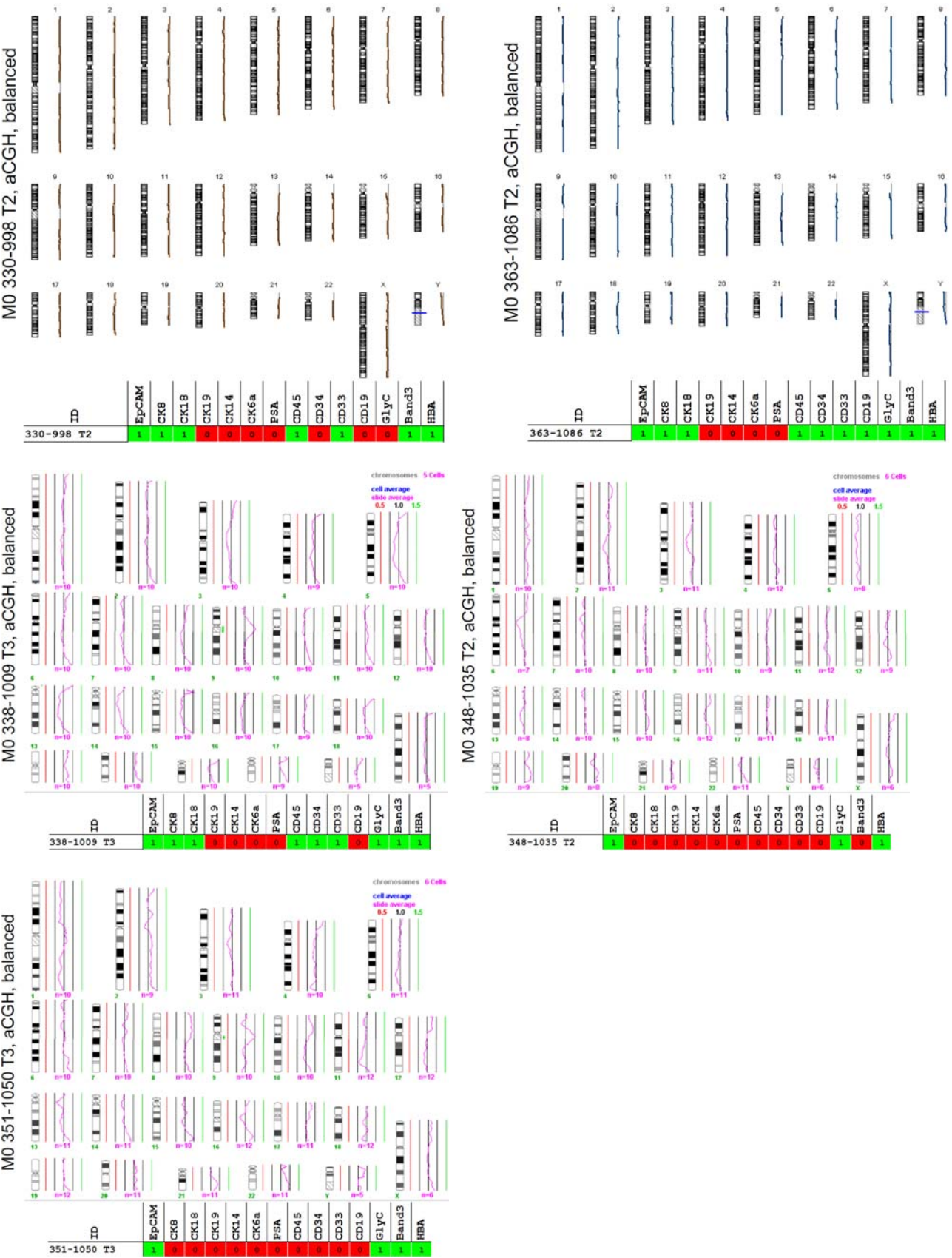
D (continued)



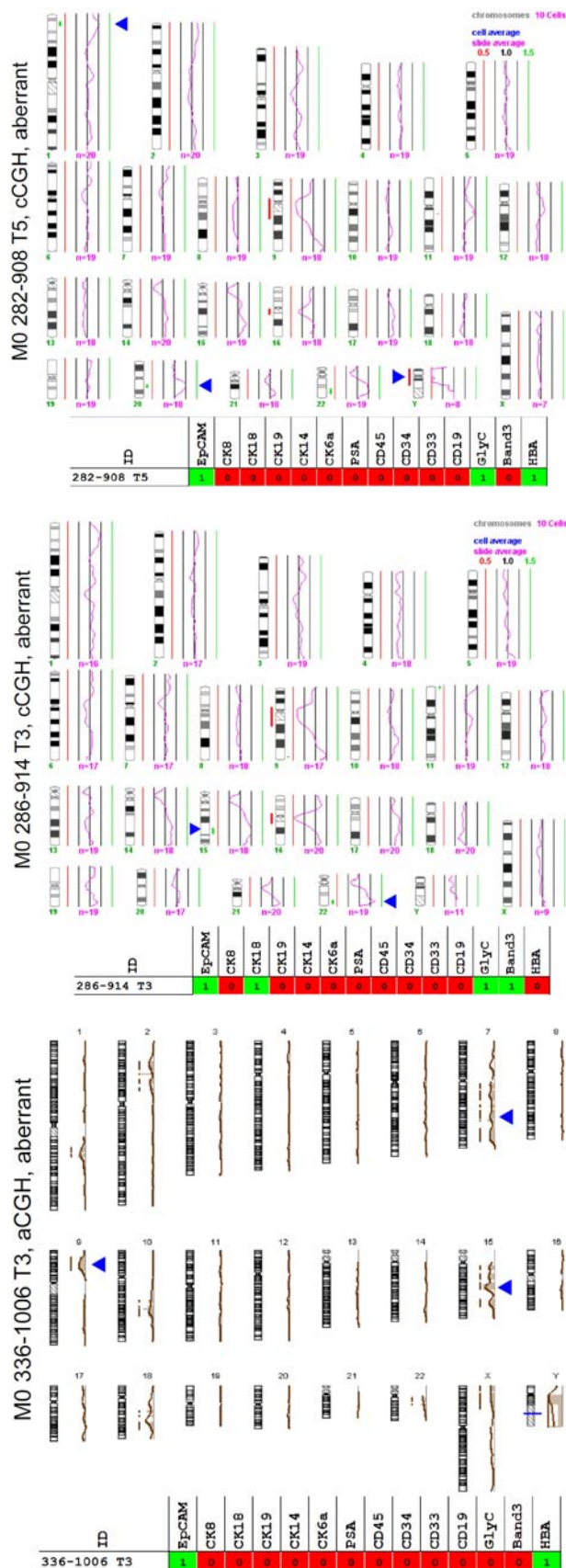


Supplementary Figure S10 (continued)

D (continued)

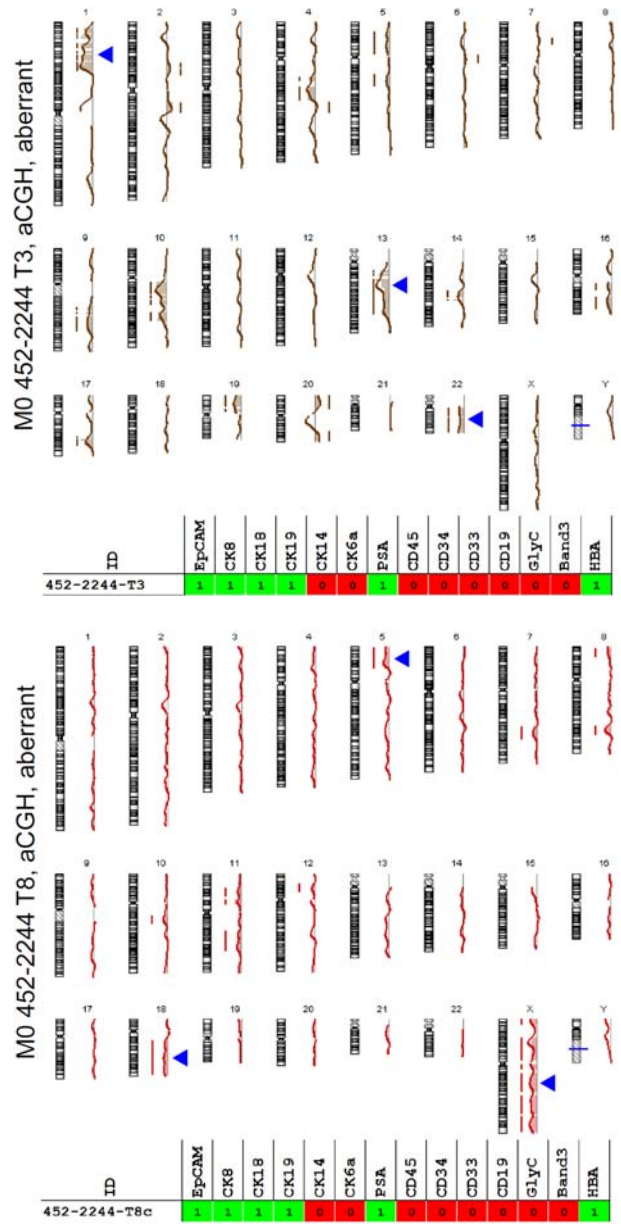
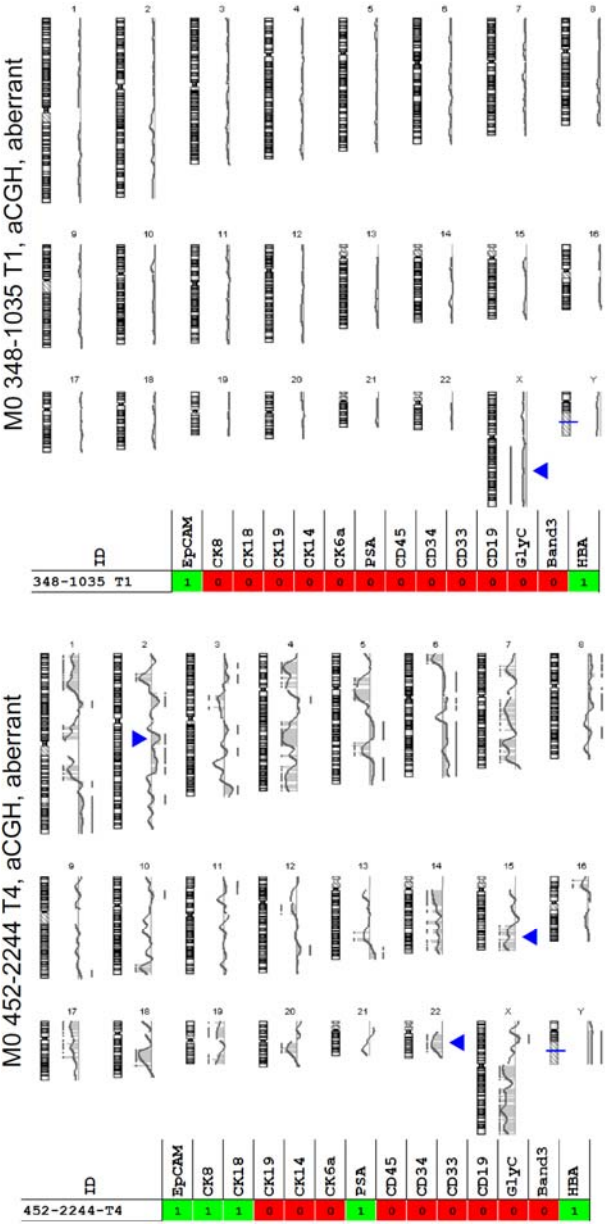


# E



Supplementary Figure S10 (continued)

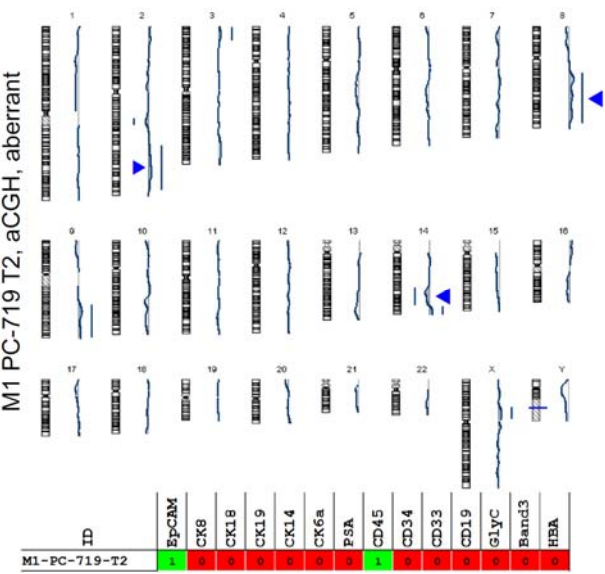
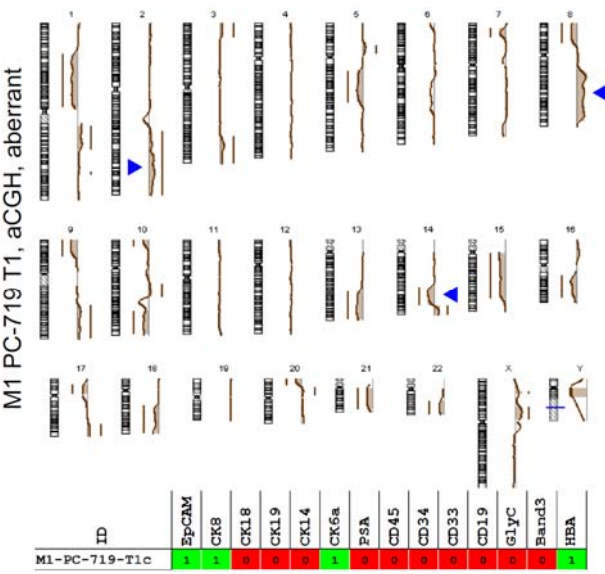
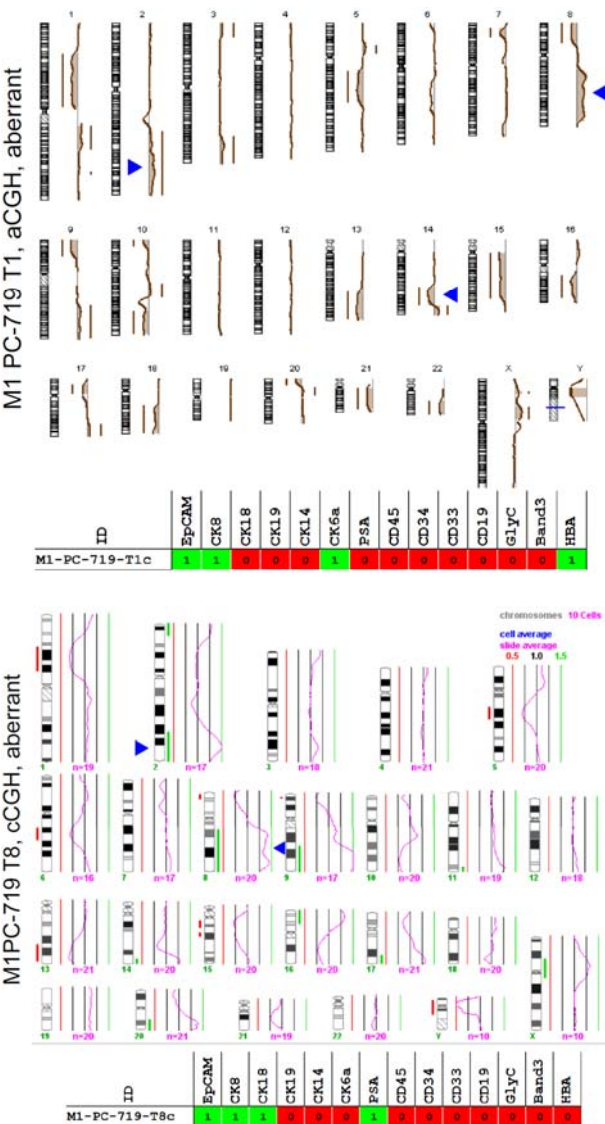
E (continued)





Supplementary Figure S10 (continued)

E (continued)





## 8. Danksagung

An erster Stelle möchte ich mich vielmals bei Herrn Prof. Dr. Christoph Klein bedanken, der mir ermöglicht hat an diesem interessanten Thema zu arbeiten, mich in all meinen Problemen unterstützt hat und mir stets volles Vertrauen entgegengebracht hat. Er war ein hervorragender Betreuer, der jederzeit präsent war und mich für meine weitere Laufbahn immer gefördert hat. Ohne seine Ideen, die fruchtbaren Diskussionen und seinen Willen sich nie mit Rückschlägen zufrieden zu geben, wäre dieses Projekt sicher nicht so weit gekommen.

Dir, lieber Mio, gilt mein ganz besonderer Dank. Du warst der beste Lehrer, den ich mir vorstellen kann. Vom ersten Moment an hast du mir zu 100 Prozent vertraut und mich sowohl gefördert als auch gefordert. Ich kann nur sagen, dass das Arbeiten mit dir stets eine große Freude für mich war: *Hvala na svemu moj prijatelju!*

Besonderer Dank auch an Bernhard Polzer, der dieses Projekt stets mit Rat und Tat begleitet hat und für alle Probleme ein offenes Ohr hatte.

Danke an alle Mitarbeiter des Lehrstuhls für Experimentelle Medizin und Therapieverfahren, die diese Arbeit unterstützt haben und mir den Laboralltag auf verschiedenste Weise erleichtert haben. Speziell möchte ich mich bei Sandra Grunewald, Isabell Blochberger, Manfred Meyer, Thomas Schamberger und Hans-Jürgen Laberer bedanken. Danke auch an Zbignies Czyz, Giancarlo Feliciello, Matthias Maneck und Martin Hoffmann für die tolle Zusammenarbeit.

Darüber hinaus möchte ich meinen Freunden und vor allem meiner Familie Danke sagen für die seelische Unterstützung während des Studiums und während der Promotion. Für ihre Liebe und Unterstützung danke ich von ganzem Herzen meinen Eltern. Ohne euch wäre all dies nicht möglich gewesen.

Zuletzt noch einen besonderen Dank an meine Regine für deine Liebe, dein grenzenloses Verständnis und deine Unterstützung bei dieser Arbeit.

## **9. Erklärung**

Ich erkläre hiermit, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet. Insbesondere habe ich nicht die entgeltliche Hilfe von Vermittlungs- bzw. Beratungsdiensten (Promotionsberater oder andere Personen) in Anspruch genommen. Niemand hat von mir unmittelbar oder mittelbar geldwerte Leistungen für Arbeit erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen. Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Regensburg, den 12.05.2015

Bernhard Braun